# Past and contemporary introgression between two strongly differentiated *Ciona* species as revealed by the analysis of post-genomic ancestry-informative SNPs

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

[1]**Sorbonne Universités, UPMC Univ Paris 6**, **CNRS,** UMR 7144, Equipe DIVCO, Station Biologique de Roscoff, Place Georges Teissier, **29680 Roscoff, France**

[2]**CNRS,** UMR 7144, Adaptation et Diversité en Milieu Marin, Station Biologique de Roscoff, Place Georges Teissier, **29680 Roscoff, France**

[3]**Université Montpellier 2**, Station Méditerranéenne de l'Environnement Littoral, 2 Rue des Chantiers, **34200 Sète, France**

[4]**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 (DIVCO), CNRS-UPMC, Station Biologique de Roscoff, Place Georges Teissier, 29680 Roscoff, France

Email: viard@sb-roscoff.fr

**Running title:** Introgression between *Ciona* species

**Keywords:** Secondary contacts, Biological invasions, Tunicates, Contemporary hybridization, Past introgression, Population genomics

# Abstract

Biological introductions bring into contact species that can still hybridize. 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. The recent secondary contact between the non-native sea squirt *Ciona robusta* (formerly known as *C. intestinalis* type A) and its native congener *C. intestinalis* (formerly known as *C. intestinalis* type B), in the western English Channel, provides an excellent case study to examine. We developed a panel of 310 ancestry-informative SNPs developed from a population transcriptomic study. Hybridization rate was examined by studying 449 individuals sampled in 8 sites from the sympatric range and 5 sites from allopatric ranges. The results clearly showed an almost complete absence of contemporary hybridization between the two species in syntopic localities, with only one first generation hybrid and no other genotype compatible with recent backcrosses. Despite lack of contemporary hybridization between the two species, shared polymorphisms were observed in sympatric and allopatric populations of both species. Furthermore, one allopatric population from SE Pacific exhibited a higher rate of introgression compared to all other *C. robusta* populations. Altogether, these results indicate that the observed level of shared polymorphism is the more probably the outcome of ancient gene flow spread afterwards at a worldwide scale. They also emphasise efficient reproductive barriers preventing hybridization between introduced and native species; this suggests that hybridization should not impede too much the expansion wave of the non-native species in its introduction range.

# 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 can deeply modify the evolution of these emerging novel species by promoting the formation of contact zones (Hewitt 2004; Maggs *et al.* 2008; Swenson & Howard 2005). In cases of species that are not fully reproductively isolated, interspecific gene flow occurs across hybrid zones (Barton 1979; Hewitt 2011). Hybridization and introgression processes between species in contact zones are particularly interesting to provide insights about the relative role of intrinsic and extrinsic barriers in the maintenance of species boundaries (Abbott *et al.* 2013; Harrison & Larson 2014; Hewitt 1988; Orr & Smith 1998; Turelli *et al.* 2001).

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.* 2016) 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.* (Currat *et al.* 2008) for theoretical supports and review of empirical evidences). And yet adaptive introgression may occur on much shorter time scale, as exemplified by introduction of species by human activities which modify species distribution at a global scale and at an unprecedented rate (e.g. in marine ecosystems see (Molnar *et al.* 2008)). Biological introductions provide a

window on the early phase 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 also reported in the literature, i.e. the rapid fixation of non-native alleles in the genome of native species, 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 reclassified although strongly differentiated species in the genus *Ciona*. These two species were considered as cryptic species of the *Ciona intestinalis* species complex and formerly named *C. intestinalis* type A and *C. intestinalis* type B (Nydam & Harrison 2007; Zhan *et al.* 2010). Following recent taxonomic revision, they are now accepted as two distinct species (WoRMS database) and respectively named *C. robusta* and *C. intestinalis* (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 ca. 15,000 years ago (95% CI: 4,300 - 56,800) during which 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 (see Supplementary Note in (Bouchemousse *et al.* 2016a)). For instance, *C. robusta*, assumed to be native to NW Pacific, have been reported as non-native species in almost all the oceans. This species live in sympatry with *C. intestinalis*, native to the NE Atlantic, in only one region, namely in the Western English Channel and South of Brittany. It has been

shown that *C. robusta* was introduced in this region probably in the early 2000s (Bishop *et al.* 2015; Nydam & Harrison 2011)).

Despite their ancient divergence, the two species are not reproductively isolated: first generation (F1) hybrids are easily obtained under laboratory conditions (Bouchemousse *et al.* 2016b; Suzuki *et al.* 2005). The question of the extent of hybridization in nature is thus to be addressed. 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) suggest contemporary hybridization happens at a small rate: despite a close syntopy and reproductive synchrony, only very few putative hybrids (i.e. individuals showing shared alleles on putative species-diagnostic markers) were observed in the wild, with a paucity of F1s (Bouchemousse *et al.* 2016b; Nydam & Harrison 2011). Given that the two taxa could have repeatedly come into secondary contacts both in past- (as inferred by (Roux *et al.* 2013)) and present time (as revealed by ecological monitoring in the English Chanel, (Bouchemousse *et al.* 2016b)), the situation is prone to the misinterpretation of contemporary admixture when few loci are used. Among the samples studied by Bouchemousse *et al.* (2016a), what appear to be a few individuals with hetero-specific alleles could indeed be a consequence of a low genome-wide level of past introgression (e.g. only 1 to 4% of the genome of present-day non-African humans derived from gene flow between Neanderthals and modern humans, (Green *et al.* 2010)). In order to conciliate the results for the previous studies, and to disentangle the relative extent of past- and contemporary inter-specific gene flow between the two species, we used a population genomic approach based on ca. 310 ancestry-informative SNPs derived from full transcriptomic sequences (Roux *et al.* 2013) and analyzed the admixture pattern of the two species sampled in eight localities of the sympatric range and two to three localities outside contact zones for each species (i.e. allopatric populations used as a control for the absence of contemporary gene flow between the two species). The SNP panel developed in

this study should prove useful in an ascidian species with importance in evolutionary biology, invasion biology, development biology and phylogeny (Procaccini *et al.* 2011; Satoh *et al.* 2014; Zhan *et al.* 2015).

# Materials and Methods

## *Sampling*

Sampling of *Ciona robusta* and *C. intestinalis* was done within their contemporary sympatric range (i.e. Western English Channel and South of Brittany) in seven localities where the two species are living in syntopy (i.e. living in the same habitat) and one locality where surveys carried out over three years never reported the presence of *C. robusta* (i.e. only *C. intestinalis* is present; no.9 in Table 1; (Bouchemousse *et al.* 2016b)). 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 (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. The assignment to either *C. robusta* or *C. intestinalis* was based both on morphological features (Brunetti *et al.* 2015; Sato *et al.* 2012) and a maternal species-diagnostic mitochondrial locus (mtCOI; (Nydam & Harrison 2007)). In addition to specimens sampled in natural populations, two F1-hybrids produced from experimental crosses (Bouchemousse *et al.* 2016b) 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) selected from a SNP dataset detected in the full transcriptomes of 10 individuals of *C. robusta* and 10 individuals of *C. intestinalis* (details in Roux *et al.* (2013)). The loci were first chosen to maximize their genotyping success: Because we used transcriptome data, we identified exons borders by aligning our data with *C. robusta* genome (vKH.71 from Ensembl; note that the genome name is misleading as it is labelled "*C. intestinalis*" although it is from *C. robusta* following the recent taxonomic revision (Brunetti *et al.* 2015)). Polymorphic sites closer than 20bp from exon border were automatically excluded. Polymorphic positions were selected within exons to produce an individual sequence for each given SNP compatible with the ADT software (available on Illumina webpage). Sites with a minor allele frequency lower than 0.1 were excluded. ADT software was used to choose primers for each SNP and estimate probability of amplification of each marker. Only markers with a probability of amplification greater than 40% were retained. The average probability obtained for our final SNP panel of 384 markers was 74%. Based on the results by Roux *et al.* (2013), loci could be sorted according to four categories of polymorphism (Table S1): 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 (note this is usual practice in hybrid zones, (Bierne *et al.* 2011)): 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 S2.

## *Intra-specific analysis*

In order to compare sympatric and allopatric populations, genetic studies were carried out for both *C. robusta* and *C. intestinalis*. Only loci that were polymorphic in the targeted species were used. For the few SNPs chosen in the same contig, only one SNP was selected, the one showing the maximum value of the minor allele frequency. Totals of 111 and 150 polymorphic loci were retained for *C. robusta* and *C. intestinalis*, respectively.

*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 v4 with default parameters for tests. *P*-values resulting of multiple tests were adjusted using the R package (R Development Core Team 2008) QVALUE (Storey 2002).

*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 Discriminant Analysis of Principal Components (DAPC;

(Jombart *et al.* 2010)) was computed for each species separately using the R package ADEGENET v.1.4 (Jombart & Ahmed 2011).

### *Genome and population admixture analysis between the two species*

To examine inter-specific gene flow at the genome level, we selected 105 loci that were the most differentiated loci according to $F_{ST}$ values ($F_{ST} > 0.9$) computed between the two species using allopatric populations (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). The hybrid indices between species and between allopatric and sympatric localities for each species were compared by permutation tests using the R package *coin* (Hothorn *et al.* 2008). 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) using the dataset of 105 loci. 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.

The inter-specific admixture rate was also investigated on the total dataset (i.e. 310 loci) using a Bayesian clustering method implemented in STRUCTURE v.2.3 (Pritchard *et al.*

2000) and a Principal Component Analysis (PCA) using the R package ADEGENET v.1.4 (Jombart & Ahmed 2011).The method implemented in STRUCTURE method used MCMC to generate posterior probabilities of assignment of each individual genotype to a number of clusters (K). Ten replicates of 150,000 MCMC after a period of 50,000 burn-in cycles were ran for a K value of 2 corresponding to the two species clusters. Results were summarized across all replicate runs using CLUMPP v.1.1.2 (Jakobsson & Rosenberg 2007) and visualized with DISTRUCT v1.1 (Rosenberg 2004).

To better evaluate the evolutionary history between *C. robusta* and *C. intestinalis* populations, we used a population graph approach implemented in the TREEMIX program (Pickrell & Pritchard 2012), which infers patterns of splitting and migration between populations. By using the matrix of allele frequency covariance between pair of populations, this method generates maximum likelihood population trees under the hypothesis of an absence of migration or the alternative hypothesis of migration event(s) (that are sequentially added). To avoid noises due small sample sizes and to intra-specific migration (i.e. infra-specific admixture), we pooled populations according to the region of sampling (i.e. no.4a and 4b for *C. intestinalis*; no.5 and 6 and no. 7 to 12 for each species). Using the total dataset (i.e. 310 loci), we search for the best tree to fit the data testing for a range of migration events from 0 to 8 (reaching an asymptotic value, Figure S3). Based on these inferences, we used a block-jackknife procedure with blocks of 10 SNPs to determine which migration events significantly improved the model fit. A complementary analysis based on $f_3$-statistic test, developed by Reich *et al.* (2009), was done to test the null hypothesis that the evolutionary history of *Ciona* populations was consistent with the absence of migration events between populations. The $f_3$-statistics evaluates the deviation of the null hypothesis using the same block-jackknife procedure for all combinations of three populations (one used as the target and two tested as putative ancestral populations).

# 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 S1. 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.

## *Population genetic structure and little heterozygosity variations in the two study species*

The analyses aiming at comparing allopatric and sympatric populations of each species were carried out separately for *C. robusta* and *C. intestinalis,* using the set of loci polymorphic in each species, i.e. 111 and 150 SNPs respectively. Results are summarized in Table 2 and Table 3.

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

Values of *He* were similar across populations of *C. robusta*, ranging from 0.234 (no.2) to 0.288 (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 S3a). The DAPC (Figure 1a) well-illustrated this finding, with a differentiation along the first discriminant axis explained by the genetic difference of the population no.1 compared to the others. The second discriminant axis pointed out the differentiation between populations of UK (i.e. no.5 and 6) and Mediterranean Sea (no.2) showing significant pairwise estimates of $F_{ST}$ (Table S3a). Populations of Brittany were relatively poorly differentiated between them (non-significant $F_{ST}$ values in most of pairwise comparisons, Table S3a). Altogether, SE Pacific and to a lesser extent UK and Mediterranean Sea 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 0.240 (no.12) to 0.229 (no.5 and 6), except for the populations from the North Sea which exhibited lower values of *He* (i.e. 0.194 and 0.172 for no.4a and 4b respectively). As for *C. robusta*, no departure from HWE was observed in any study populations. Exact test of differentiation between *C. intestinalis* populations indicated significant differences among all populations but was non-significant 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 almost all other populations (pairwise comparisons are provided in Table S3b). These patterns are pictured by the DAPC (Fig.1b).

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

A total of 105 loci, showing a $F_{ST}$ strictly superior to 0.9, were used to examine the patterns of shared polymorphism between the two species, and the likelihood of contemporary

inter-specific hybridization vs. past introgression events. At the species level (i.e. across all individuals for each species), values of *h* were very low, with an average value across individuals of 0.0029 for *C. robusta* and 0.0055 for *C. intestinalis*. *h* was however significantly higher for individuals of *C. intestinalis* than *C. robusta* (test per permutation on *h* value of individuals, *Z-value* = 6.172, $P < 0.001$). 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 0.001 for the allopatric population of Mediterranean Sea (Etang de Thau, no.2) to 0.056 for the allopatric population of SE Pacific (Guanaqueros, no.1). Populations sampled in the sympatric range showed intermediate *h*-values, from 0.002 (no.11) to 0.004 (no.8). The difference of *h* was however non-significant between populations from syntopic and non-syntopic localities (*Z-value* = 1.005, $P = 0.302$). Conversely, the large value of *h* in the allopatric population no.1 was significantly higher than almost all others populations sampled (pairwise comparison in Table S3a); while (except for one comparison), *h* values of other populations were not significantly different between them (Table S4a). For *C. intestinalis*, the average of *h* across individuals per population ranged from 0.001 for the allopatric population of North Sea (Fiskebackskil – 20m depth, no.4b) to 0.029 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 0.5. This individual was assigned with a probability of 1 to a 'F1 hybrid' with NEWHYBRIDS (Table 1). When removing this individual from the *h* estimation, the value in Camaret dropped to 0.006, a value close to the average values for *C. intestinalis* individuals (i.e. 0.004). It is noteworthy that all the other study individuals were assigned to their respective parental 'species' categories (Table 1). Significant lower values of *h* was found in populations of non-syntopic localities compared to populations of syntopic

localities (*Z-value* = 1.987, *P* = 0.036), a result mainly explained by the value of *h* in population no.4b which was significantly lower than almost all of the others populations (without no.4b, *Z-value* = 1.987, *P* = 0.061, pairwise comparison in Table S4b). When analyzing syntopic localities, no significant correlation was observed for *h* between the populations of the two species sampled in these localities (Pearson correlation, $r^2$ = -0.407, *P* = 0.123). The relationship between *h*-value and the heterozygosity rate across the 105 study loci was plotted using a triangle plot displayed in Figure 2: 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. 99%); these values were similar to the values observed for the two F1-hybrids experimentally produced in the laboratory (Fig.2, i.e. 96% and 99%). This hybrid assignation was also well-highlighted by using all the markers available (i.e. 310 loci): STRUCTURE analyses assigned equally the putative wild F1-individual to the two species clusters (Fig.3b); and the results of the PCA showed a clear distribution of the overall genetic variance between the two study species with the putative F1-individual and the two experimental F1-hybrids at an intermediate position (Fig.3a).

## *Heterogeneous polymorphism rates at the genome level*

The subset of 105 SNPs used for examining hybridization and introgression patterns (i.e. showing $F_{ST}$ values higher than 0.9) showed the following patterns over the whole dataset: 65 were differentially fixed (i.e. sf loci), 39 with private polymorphisms (i.e. sxA or sxB but with extremely low frequency of the minor allele; 8 polymorphic in *C. robusta* and 31 in *C. intestinalis*) and only one locus showing shared polymorphism (i.e. ss locus) between populations of the two species (snp18 on the chromosome 1). The 40 loci showing shared and

private polymorphisms were distributed everywhere along the genome of the two species (Fig.4a), as previously observed in Roux *et al.* (2013).

Among the 105 SNPs, some were found in two introgression hotspots defined by Roux et al. (2013). Note that we did not have SNPs localized in the other two introgression hotspots. Interestingly, we found shared or polymorphic SNP in these introgression hotspots: 1) one SNP in the introgression hotspot of chromosome 1 showed shared polymorphism in the two species, and 2) six loci showed private polymorphism in one or the other of the two species in the introgression hotspot on the chromosome 2.

Polymorphism patterns were also informative regarding the status of the study populations (i.e. allopatric and sympatric). Details of allele frequencies at each of the 40 polymorphic loci in the allopatric and sympatric ranges of the two species are provided in Table S5. When comparing of populations for *C. robusta* and *C. intestinalis*, the rates of shared and private polymorphism appeared to be remarkably stable across populations: for the two species, individuals of each population carried a small number of heterozygous sites, but not always at the same genome location (Fig.4a). A noteworthy exception was the allopatric population of *C. robusta* from Chile (no.1) which shared for some loci more polymorphism with *C. intestinalis* populations than with other populations of *C. robusta* (Fig.4a): heterozygous sites were more important, for example, at the snp18 (chromosome 1), snp290 (chrom. 2) and snp237 (chrom. 10), the two first being in introgression hotspots defined by Roux *et al.* (2013). This finding was already visible in the results of the PCA (Fig.3a) as the population of Chile was slightly shifted towards the *C. intestinalis* points.

### *Admixture events between the two species revealed by a population tree approach*

The population tree inferred from TREEMIX without migration explained 88.5% of the variance in the population covariance matrix. Note that in the population tree without migration events, the population of Chile (no.1) showed a position shifted towards *C. intestinalis* populations (Figure S4). The variance explained was increased when migration events were added (Figure S3). The best fit to the data was obtained with two migration events, which significantly improved the model ($P<0.001$, Fig.5). This population tree, explaining 98.5% of the variance (Figure S3), indicated significant gene flow in *C. robusta* population of Chile (no.1) and in the *C. intestinalis* populations group of Brittany (no.7 to 12). These two migration events were also supported by the $f_3$ statistics analysis (Table S6) with significant negative values for almost all combinations of three populations involving as targets the *C. robusta* population of Chile (no.1) and the *C. intestinalis* populations group of Brittany (no.7 to 12). $f_3$ statistics also showed significant negative values for combinations of three populations involving as targets *C. robusta* populations groups of UK (no.5 and 6) and Brittany (no.7 to 12) (Table S6). These negative $f_3$ statistics are consistent with the hypothesis that the tested populations were the results of admixture with ancestors in the two tested population sources (Reich *et al.* 2009).

# Discussion

In this study we used 310 ancestry-informative SNPs to clarify relative contribution of contemporary hybridization *versus* past introgression in the level of shared polymorphism observed between *Ciona robusta* and *Ciona intestinalis*, and to analyze the introgression patterns within allopatric and sympatric ranges of the two species. These two points are discussed in turn below.

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

In previous studies that analysed interspecific gene flow in the sympatric range, admixture have been observed between the two species although at low rates: 4.2% (including one putative F1-hybrid) over 730 individuals sampled in Nydam & Harrison (2011), 6.3% over 288 individual sampled in one locality by Sato *et al.* (2014), 4.3% (including one putative F1 hybrid) over ca. 3,000 individuals by Bouchemousse *et al.* (2016b). These authors used only few nuclear markers, supposed to be and used as being species-diagnostic (between 3 and 6 loci according to the study). Consequently, discriminating the footprint left by historical introgression or contemporary hybridization was particularly difficult as acknowledged in these studies (Bouchemousse *et al.* 2016b; Nydam & Harrison 2011). Using a large number of informative loci with a large sample, we found that with the exception of a single individual all the other occurrence of shared polymorphism are the likely consequence of a low level of past introgression, i.e. ancient gene flow between the two species, rather than contemporary hybridization. At a given locus some individuals can be found heterozygotes at quasi diagnostic loci, but averaging at many such loci shows every individual have the same hybrid index value. The only contemporary hybrid was a F1, as supported by both NEWHYBRIDS and INTROGRESS analyses (Table 1, Fig.2 and 4a). The mtDNA type of this individual is typical of *C. intestinalis* and in many studies (Bouchemousse *et al.* 2016b; Suzuki *et al.* 2005), 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.* 2016b)). The presence of one F1 hybrid only in our study confirms the hypothesis by Sato *et al.* (2014) and Bouchemousse *et al.* (2016b) of the existence of pre-zygotic isolation mechanisms preventing contemporary hybridization in the wild. However, our new interpretation that recent backcrosses of a few generations are completely lacking from the sympatric range, suggest that strong post-zygotic selection is also occurring, which is

the least one can expect for two highly divergent species (i.e. 14% of divergence based on transcriptomic data (Roux *et al.* 2013)). Dobzhansky-Muller incompatibilities expressed by recessive mutations in subsequent generations of hybridization (e.g. (Bierne *et al.* 2006; Fishman & Willis 2001) and see for a review (Maheshwari & Barbash 2011)) are likely the cause of this isolation. Altogether these results confirm that no contemporary gene flow has occurred between the two species.

### *The footprint 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). Signatures for shared polymorphism were observed in all populations including in localities of allopatric regions (Fig.4a) and was significant for some of them according to TREEMIX and $f_3$-statistics analyses. With one exception in SE Pacific, the admixture profile is also remarkably stable across populations sampled.

For a given locus, shared polymorphism or high derived allele frequency between two species may result from incomplete lineage sorting of ancestral polymorphism, contemporary or past secondary introgression, or homoplasy. In the case of the two *Ciona* species studied here, the contemporary introgression hypothesis can be reasonably excluded as discussed above. Concerning incomplete lineage sorting of ancestral polymorphism, it would have meant 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). High effective population sizes moderates the effect of genetic drift and so the probability of fixation of

alleles over the time (Maddison 1997; Pamilo & Nei 1988). *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. However, the analysis of Roux *et al.* (2013) showed that the strong excess of shared polymorphism between the two species cannot be obtained without secondary introgression. The secondary contact has been estimated to have occurred 15,500 years ago (95% CI: 4,300-56,800), during which ca. 20% of loci crossed the species barrier in both directions. Besides similarities in admixture levels across localities, the hypothesis of an ancient admixture event is also well-supported 1) by significant admixture events between populations of *C. robusta* and *C. intestinalis* according to TREEMIX and $f_3$-statistics analyses and 2) the presence of admixed loci in introgression hotspots (i.e. loci pointed by an asterisk in Fig.4a).

Our finding is also interesting to consider in light of previous studies (e.g. (Bouchemousse *et al.* 2016b; Nydam & Harrison 2011; Sato *et al.* 2014)). Analyzing a small number of such loci can easily result in the erroneous interpretation that some individuals are more admixed than other and cast doubts about the ability of these markers to reliably distinguish the two species. To better investigate the properties of these markers, we genotyped most of the individuals of the present study using three of these putative diagnostic loci; namely Hox5 (Caputi *et al.* 2007) vAChTP and CesA (Nydam & Harrison 2010). Patterns of admixture in natural populations of the two species are displayed in Fig.4b 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). However, their introgression rate is low: minor allele frequency observed for Hox5, vAChTP and CesA is 0.2, 0.2 and 1.3% for *C. intestinalis*, and

6.4, 1.2 and 0.5% for *C. robusta*, respectively. This explains why they often appear as diagnostic markers.

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).

### *Difference of introgression rate in Chile caused by adaptive introgression?*

Admixture profiles were remarkably stable across populations of allopatric and sympatric ranges. 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 contact between the two taxa, during which interspecific gene flow occurred. Genetic differentiations are however reported between allopatric and sympatric populations for the two species (Table 3) suggesting that intraspecific divergence history for each species influence more the genetic differentiation between populations than different rates of introgression between species. For example, the two sub-populations of North Sea (i.e. no.4a and 4b sampled both at Fiskebackskil at the surface and at 20m depth, respectively) exhibited a strong genetic differentiation with the other *C. intestinalis* populations and also between them (Fig.1b, Table S3) while they showed similar hybrid index values (Table S4). This strong genetic differentiation could be explained by a reduced gene flow between the two sub-populations in North Sea, a result which echoed to the pattern described in the doctoral thesis of Elin

Renborg (https://gupea.ub.gu.se/handle/2077/35128). The poor connectivity is hypothesized to result of density discontinuity of sea water which separates shallow and deep populations of *C. intestinalis*. Such patterns of population differentiation have already been documented in other coastal marine species showing extended distribution along depth gradient (Jennings *et al.* 2013; Pivotto *et al.* 2015).

A noteworthy exception of the stability of admixture profiles is the *C. robusta* population from Chile which shared more polymorphism with *C. intestinalis* than other populations. This *C. robusta* population showed the highest number of loci with polymorphism shared with *C. intestinalis* (Fig.4a) and the highest *h*-values over all *C. robusta* populations (Table 1). Moreover, the position of the Chilean population on the first axis of the PCA (Fig.2a) first suggests residual genotypic covariance best explained by a higher level of introgression by *C. intestinalis* than other *C. robusta* populations. This is formally tested using TREEMIX and $f_3$ statistical analyses (Fig.5, Table S6) which highlighted significant migration events between *C. intestinalis* ancestor and the Chilean population. Incomplete lineage sorting of ancestral polymorphism is not expected to create such asymmetry of shared polymorphism between populations, but point out evidence of local introgression in the Chilean population (Fraisse *et al.* 2016; Martin *et al.* 2013; Pickrell & Pritchard 2012). This pattern of local introgression is not uniformly distributed among loci (Fig.4a), which is usually not accounted for in demographic inferences such as TREEMIX and could explain why the source of admixture is not a contemporary *C. intestinalis* population. This pattern could be a consequence of adaptive introgression in the genomic region of these introgressed loci, a process documented in several recent studies (Fontaine *et al.* 2015; Mendez *et al.* 2012; Pardo-Diaz *et al.* 2012). A similar pattern was observed in the *Mytilus* mussel complex of species where local introgression proved to be heterogeneous across loci (Fraisse *et al.* 2016). However, other processes can generate heterogeneous introgression rates such as

heterogeneous load of deleterious mutations in migrant tracks (Christe *et al.* 2016; Harris & Nielsen 2016). None of the loci identified matched with genes coding for a known phenotypic or physiological trait (Table S5).

It is important to note that the first report of *C. robusta* along the Chilean coasts (with the name of *C. intestinalis* used until the recognition of *C. robusta* as a valid species) dates back to the middle of the 20$^{th}$ century (Van Name 1945). We thus cannot exclude that local 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). A recent phylogeographic study based on mtDNA data (Bouchemousse *et al.* 2016a) pointed out a low genetic differentiation between populations of Chile and populations sampled in Japan, the putative native range of *C. robusta*. Further analyses are needed to investigate if this pattern could be due to adaptive introgression, using for instance modelling methods such as those performed by Fraisse *et al.* (2014) in a *Mytilus* sp. hybrid zone to examine the likelihood of adaptive introgression. A much larger number of population representatives of the global distribution of *C. robusta*, particularly populations of the Asian range, 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 sympatry/syntopy. Efficient reproductive barriers seem to prevent hybridization in the wild between the two species. These results are casting doubts that hybridization could impede the spread of the non-native. Ecological processes (e.g. niche displacement, trophic competition) might thus be more important to determine the fate of the two species in the sympatric range. Even if efficient reproductive isolation mechanisms are acting, few crosses involving an

advantageous allele can be sufficient to favor its transmission in subsequent generations of the non-native species (Hedrick 2013). Our density of markers was clearly not sufficient to detect local signatures of adaptive introgression at genomic level. High-throughput genome analyses will be needed to definitively exclude, or confirm, that invasion potential of *C. robusta* is facilitated by adaptive introgression with *C. intestinalis* in the Northeast Atlantic. Altogether, our study provides evidence that what was inferred to be backcrossed individuals are more likely the outcome of a low level of residual historical introgression redistributed at global scale by natural range shifts and human-mediated introductions. Local introgression patterns, mostly concentrated on a few genome regions, were observed in the population sampled in the SE Pacific, a population far from the current distribution range of *C. intestinalis*. This result paves the way for further work to investigate adaptive introgression processes in other regions, in light of the range shift history of *C. robusta*.

# Acknowledgment

The authors are very grateful to C. Roux and N. Galtier for making available RNA sequences from the Pophyl Project, K. Belkhir for his precious help to the optimization of loci selection and the ADN$^{id}$ society (Montpellier) for the genotyping of SNPs. We are also grateful to all our colleges who contributed to the collection of samples: the divers of the Marine Operations department (*Service Mer & Observation*) at the Roscoff Biological Station, J.D.D. Bishop, S. Krueger-Hadfield, B. Lundve, J. Pechenik. The authors kindly acknowledge C. Roux and C. Fraisse for help and advices on R packages and scripts. This work benefitted from funding of the ANR project HYSEA (no. ANR-12-BSV7-0011) and the Interreg IVa Marinexus project.

# Data accessibility

Full dataset of the 310 SNPs were deposited into the DRYAD database (DOI: XXXXX).

## Author contributions

SB, CHL, NB and FV designed the study. CHL and NB performed the choice of SNP panel. SB and FV performed the choice of populations for genotyping. SB, CHL, NB and FV analyzed the data and wrote the article.

# References


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

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.

Bishop JDD, Wood CA, Yunnie ALE, Griffiths CA (2015) Unheralded arrivals: non-native sessile invertebrates in marinas on the English coast. *Aquatic Invasions* **10**, 249-264.

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, Bishop JDD, Viard F (2016a) Contrasting global genetic patterns in two biologically similar, widespread and invasive *Ciona* species (Tunicata, Ascidiacea). *Scientific Reports* **6**, 24875.

Bouchemousse S, Lévêque L, Dubois G, Viard F (2016b) Co-occurrence and reproductive synchrony do not ensure hybridization between an alien tunicate and its interfertile native congener. *Evolutionary Ecology* **30**, 69-87.

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.

Christe C, Stölting KN, Bresadola L, *et al.* (2016) Selection against recombinant hybrids maintains reproductive isolation in hybridizing *Populus* species despite F1 fertility and recurrent gene flow. *Molecular Ecology* **25**, 2482-2498.

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

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.

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.

Fontaine MC, Pease JB, Steele A, *et al.* (2015) Extensive introgression in a malaria vector species complex revealed by phylogenomics. *Science* **347** (6217), 1258524.

Fraisse C, Belkhir K, Welch J, Bierne N (2016) Local interspecies introgression is the main cause of extreme levels of intraspecific differentiation in mussels. *Molecular Ecology* **25**, 269-286.

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.

Green RE, Krause J, Briggs AW, *et al.* (2010) A draft sequence of the Neandertal genome. *Science* **328**, 710-722.

Harris K, Nielsen R (2016) The genetic cost of Neanderthal introgression. *Genetics* **203**, 881-891.

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.

Hewitt GM (2011) Quaternary phylogeography: the roots of hybrid zones. *Genetica* **139**, 617-638.

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.

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** (10), e77594.

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

Jombart T, Devillard S, Balloux F (2010) Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. *Bmc Genetics* **11**: 94.

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**, 108-122.

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.

Martin SH, Dasmahapatra KK, Nadeau NJ, *et al.* (2013) Genome-wide evidence for speciation with gene flow in *Heliconius* butterflies. *Genome Research* **23**, 1817-1828.

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.

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**(6), e1002752.

Pickrell JK, Pritchard JK (2012) Inference of population splits and mixtures from genome-wide allele frequency data. *Plos Genetics* **8**(11), e1002967.

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.

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 P), pp. 91-106.

R Development Core Team (2008) R: A language and environment for statistical computing. *R Fondation for Statistical Computing*, Vienne, Austria. ISBN: 3-900051-07-0. http://www.R-project.org.

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

Reich D, Thangaraj K, Patterson N, Price AL, Singh L (2009) Reconstructing Indian population history. *Nature* **461**, 489-494.

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

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

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.

Storey J (2002) A direct approach to false directory rates. *Journal of the Royal Statistical Society, Series B* **64**, 479-498.

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.

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

Van Name WG (1945) The North and South American ascidians. *Bulletin of the American Museum of Natural History* **84**, 1-476 pls. 471-431.

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

Zhan A, Briski E, Bock DG, Ghabooli S, MacIsaac HJ (2015) Ascidians as models for studying invasion success. *Marine Biology* **162**, 2449-2470.

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 according to NEWHYBRID software) **in each population of *Ciona robusta* and *C. intestinalis*.**

| N° | Locality | Region | Coordinates (Long., Lat.) | Regional status | Locality status | Sampling year | Nind | Hybrid index (mean ± SD) | Number of hybrids |
|---|---|---|---|---|---|---|---|---|---|
| ***C. robusta*** | | | | | | | | | |
| 1- | Guanaqueros, Ch | South Eastern Pacific | -30.1945, -71.4300 | Allopatric | Non-syntopic | 2012 | 24 | 0.0056 ± 0.0039 | 0 |
| 2- | Etang de Thau, Fr | Mediterranean Sea | 43.4014, 3.6582 | Allopatric | Non-syntopic | 2013 | 23 | 0.0010 ± 0.0020 | 0 |
| 5- | Falmouth, UK | English Channel | 50.1543, -5.0579 | Sympatric | Syntopic | 2013 | 24 | 0.0024 ± 0.0031 | 0 |
| 6- | Plymouth, UK | English Channel | 50.3583, -4.1228 | Sympatric | Syntopic | 2011 | 24 | 0.0024 ± 0.0028 | 0 |
| 7- | St Vaast, Fr | English Channel | 49.5897, -1.2648 | Sympatric | Syntopic | 2012 | 23 | 0.0031 ± 0.0034 | 0 |
| 8- | Perros Guirec, Fr | English Channel | 48.8112, -3.4295 | Sympatric | Syntopic | 2011 | 24 | 0.0038 ± 0.0037 | 0 |
| 10- | Moulin blanc, Fr | Bay of Brest | 48.3906, -4.4318 | Sympatric | Syntopic | 2012 | 24 | 0.0034 ± 0.0041 | 0 |
| 11- | Camaret, Fr | Bay of Brest | 48.2799, -4.5961 | Sympatric | Syntopic | 2011, 2012 | 24 | 0.0020 ± 0.0031 | 0 |
| 12- | Quiberon, Fr | Bay of Biscay | 47.4858, -3.0999 | Sympatric | Syntopic | 2012, 2013 | 23 | 0.0025 ± 0.0024 | 0 |
| | **Total** | | | | | | **213** | **0.0029 ± 0.0034** | |
| ***C. intestinalis*** | | | | | | | | | |
| 3- | Nahant, USA | North Western Atlantic | 42.4569, -70.9414 | Allopatric | Non-syntopic | 2013 | 24 | 0.0064 ± 0.0056 | 0 |
| 4a- | Fiskebackskil – surface, Sw | North Sea | 58.2502, 11.4579 | Allopatric | Non-syntopic | 2010 | 12 | 0.0016 ± 0.0024 | 0 |
| 4b- | Fiskebackskil - 20m depth, Sw | North Sea | 58.2502, 11.4579 | Allopatric | Non-syntopic | 2010 | 12 | 0.0012 ± 0.0022 | 0 |
| 5- | Falmouth, UK | English Channel | 50.1543, -5.0579 | Sympatric | Syntopic | 2011 | 24 | 0.0042 ± 0.0043 | 0 |
| 6- | Plymouth, UK | English Channel | 50.3583, -4.1228 | Sympatric | Syntopic | 2011 | 24 | 0.0060 ± 0.0058 | 0 |
| 7- | St Vaast, Fr | English Channel | 49.5897, -1.2648 | Sympatric | Syntopic | 2012 | 23 | 0.0081 ± 0.0075 | 0 |
| 8- | Perros Guirec, Fr | English Channel | 48.8112, -3.4295 | Sympatric | Syntopic | 2011, 2012 | 24 | 0.0070 ± 0.0069 | 0 |
| 9- | Aber Wrac'h, Fr | English Channel | 48.5987, -4.5622 | Sympatric | Non-syntopic | 2011, 2012 | 23 | 0.0064 ± 0.0047 | 0 |
| 10- | Moulin blanc, Fr | Bay of Brest | 48.3906, -4.4318 | Sympatric | Syntopic | 2011 | 24 | 0.0071 ± 0.0056 | 0 |
| 11- | Camaret, Fr | Bay of Brest | 48.2799, -4.5961 | Sympatric | Syntopic | 2011, 2012 | 22 | 0.0286 ± 0.1065 | **1 (F1-hybrid)** |
| | *without F1 -hybrid* | | | | | | *21* | *0.0059 ± 0.0052* | *0* |
| 12- | Quiberon, Fr | Bay of Biscay | 47.4858, -3.0999 | Sympatric | Syntopic | 2011, 2012 | 24 | 0.0062 ± 0.0065 | 0 |
| | **Total** | | | | | | **236** | **0.0055 ± 0.0241** | **1 (F1-hybrid)** |
| | ***Total (without F1-hybrid)*** | | | | | | ***235*** | ***0.0044 ± 0.0050*** | ***0*** |

*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). In this table, *h* is defined as the proportion of alleles from one species in the genetic background of the other species (i.e. proportion of *C. intestinalis* alleles over all loci in *C. robusta* individuals and proportion of *C. robusta* alleles over all loci in *C. intestinalis* individuals). *h* values were averaged across individuals for each sampled localities. Analyses were done with 105 SNPs selected for inter-specific gene flow analyses ($F_{ST}$ > 0.9; see Material and Methods).

Table 2. **Genetic diversity indices and fixation index** of each study populations for *Ciona robusta* and *C. intestinalis*.

| N° | Locality | Introduced vs. native status | $P_{loc}$ | $H_e$ | $F_{IS}$ |
|---|---|---|---|---|---|
| ***C. robusta*** | | | | | |
| 1- | Guanaqueros | Introduced | 103 | 0.288 | -0.013 |
| 2- | Etang de Thau | Introduced | 77 | 0.234 | -0.012 |
| 5- | Falmouth | Introduced | 79 | 0.247 | -0.055 |
| 6- | Plymouth | Introduced | 80 | 0.240 | -0.043 |
| 7- | St Vaast | Introduced | 79 | 0.238 | -0.076 |
| 8- | Perros Guirec | Introduced | 81 | 0.246 | -0.024 |
| 10- | Moulin blanc | Introduced | 81 | 0.236 | -0.043 |
| 11- | Camaret | Introduced | 81 | 0.247 | -0.022 |
| 12- | Quiberon | Introduced | 83 | 0.254 | -0.029 |
| | ***Total (Sympatric pop.)*** | | ***86*** | ***0.253*** | |
| | **Total** | | **111** | **0.265** | |
| ***C. intestinalis*** | | | | | |
| 3- | Nahant | Cryptogenic | 117 | 0.233 | -0.018 |
| 4a- | Fiskebackskil - surface | Native | 97 | 0.194 | 0.048 |
| 4b- | Fiskebackskil - 20m depth | | 85 | 0.172 | 0.001 |
| 5- | Falmouth | Native | 118 | 0.229 | -0.012 |
| 6- | Plymouth | Native | 125 | 0.229 | 0.002 |
| 7- | St Vaast | Native | 123 | 0.233 | 0.032 |
| 8- | Perros Guirec | Native | 125 | 0.234 | -0.035 |
| 9- | Aber Wrac'h | Native | 122 | 0.239 | -0.005 |
| 10- | Moulin blanc | Native | 125 | 0.239 | 0.014 |
| 11- | Camaret | Native | 120 | 0.234 | -0.011 |
| 12- | Quiberon | Native | 120 | 0.240 | -0.014 |
| | ***Total (Sympatric pop.)*** | | ***148*** | ***0.242*** | |
| | **Total** | | **150** | **0.244** | |

$P_{loc}$: number of polymorphic loci and $H_e$: expected heterozygosity over 111 and 150 polymorphic loci retained for intra-specific analyses in *C. robusta* and *C. intestinalis*, respectively (see *Materials and Methods*); $F_{IS}$: fixation index calculated (no deviation from Hardy-Weinberg equilibrium; exact test, $P < 0.05$).

Table 3. **Genetic structure among populations for *Ciona robusta* and *C. intestinalis*.**

| | $F_{ST}$ | *P*-value |
|---|---|---|
| ***C. robusta*** | | |
| All sampled populations (9 populations) | 0.054 | $P < 0.001$ |
| 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) | 0.045 | $P < 0.001$ |
| All populations without Fiskebackskil (all except no.4a and 4b) | 0.021 | $P < 0.001$ |
| Sympatric populations (all except no.3, 4a and 4b) | 0.014 | $P = 0.020$ |

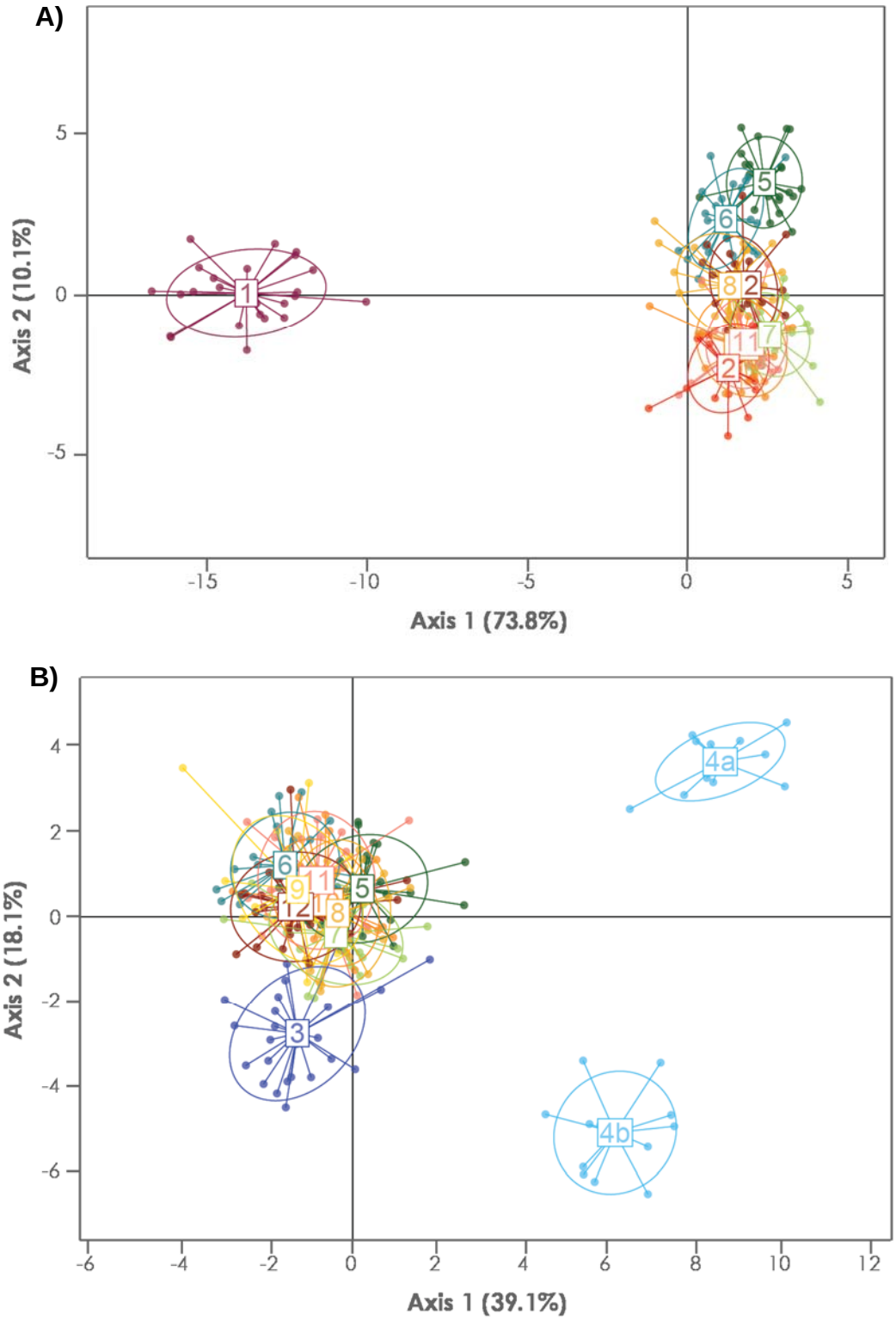


Figure 1. **Discriminant Analysis of Principal Components (DAPC) among populations of *C. robusta* (A) and *C. intestinalis* (B).** Only the two first axis showing the two higher discriminant eigenvalues are presented here.

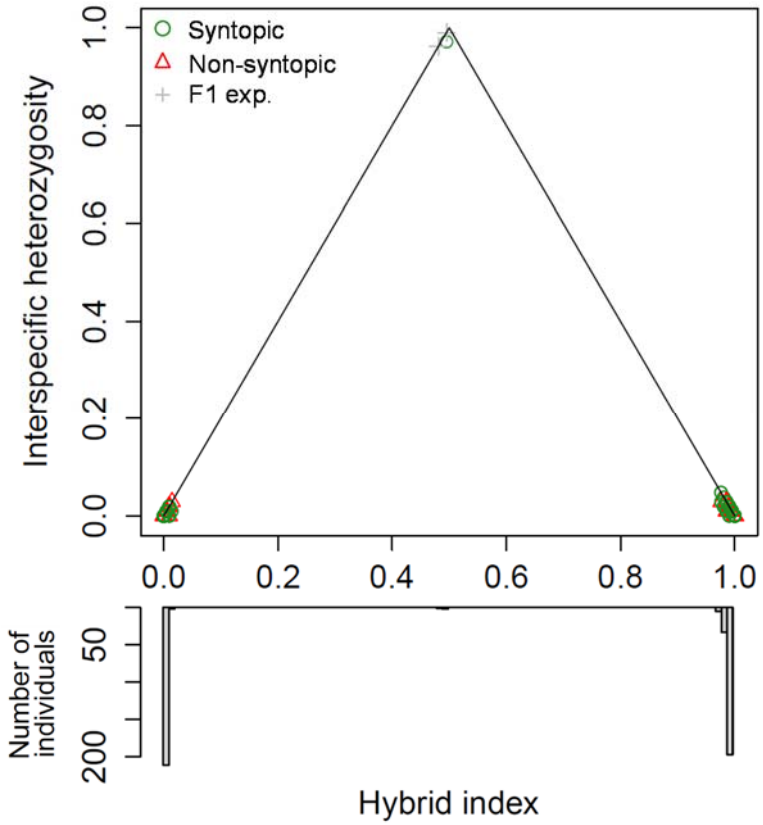


Figure 2. **Triangle plot showing the relationship between heterozygosity rate across loci and hybrid index for each individual**. At the top of the triangle, one green circle is picturing one individual from the locality no. 11 and the two gray crosses are F1-hybrids from experimental crosses.

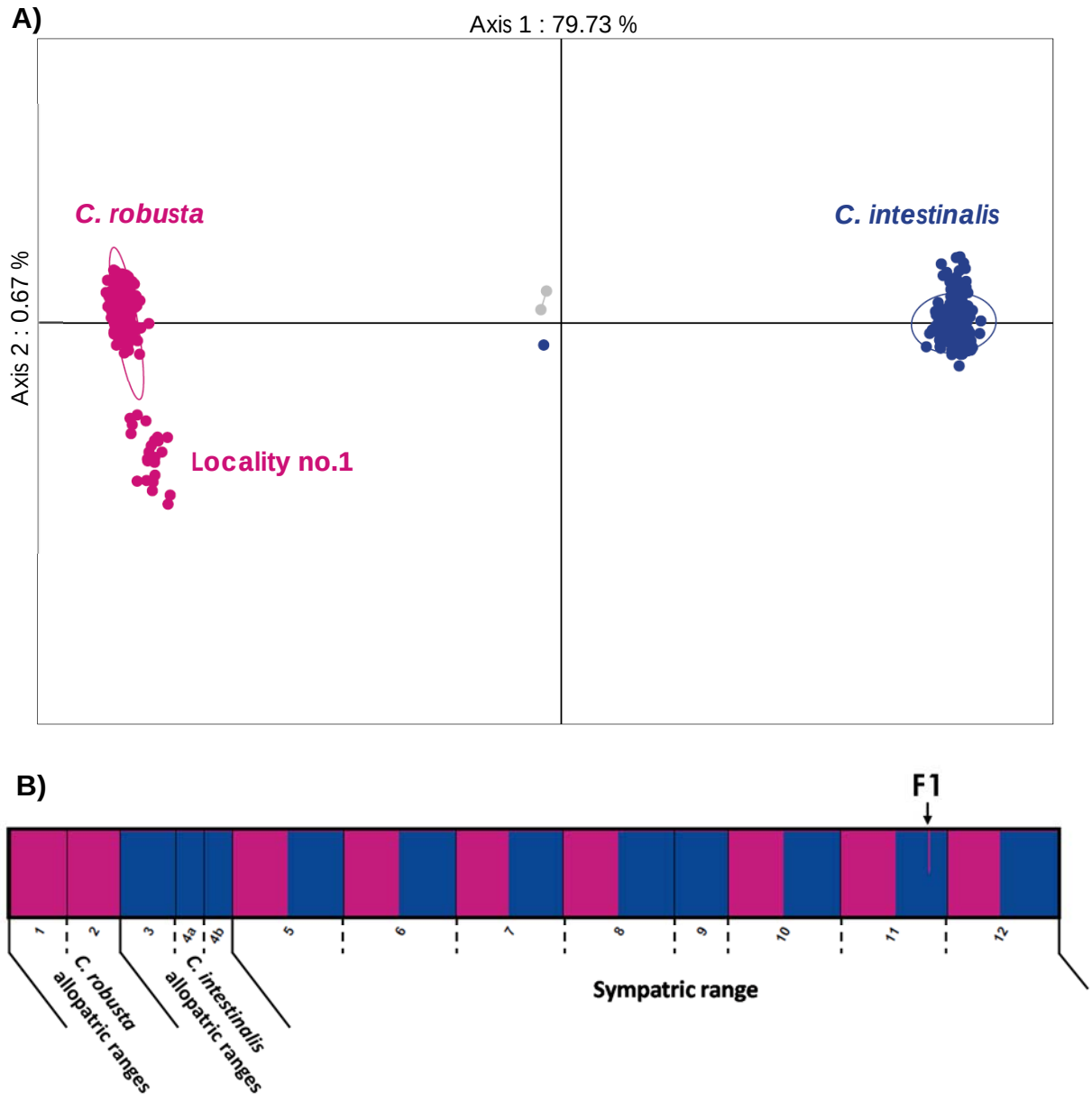


Figure 3. **A) Principal Components Analysis (PCA) and B) Individual Bayesian assignment proportion for K=2 clusters** using the total dataset (i.e. 310 SNPs, 449 individuals from natural population). Note that the two F1 hybrids from experimental crosses were added in the PCA (grey points).

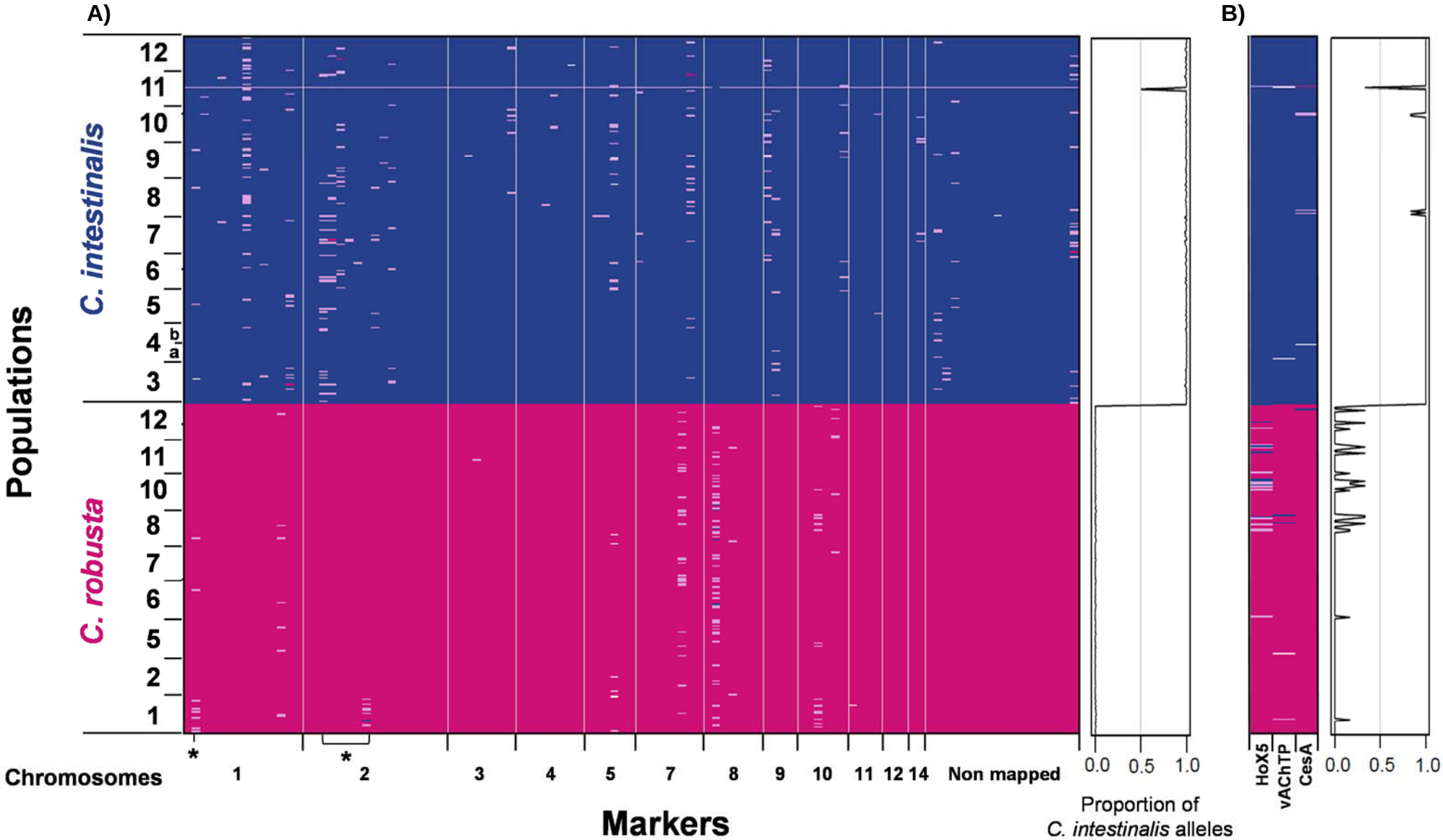


Figure 4. **A) Genomic architecture using 105 highly differentiated loci ($F_{st}$ > 0.9) selected for inter-specific analyses**. Markers (x-axis) are ordered following physical position on chromosomes. Individuals (y-axis) are ordered per population. Dark pink cases indicate homozygote genotype on *C. robusta* alleles; dark blue, homozygote genotype on *C. intestinalis* alleles; light purple, heterozygotes for *C. robusta* and *C. intestinalis* alleles; and white cases, missing values. 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.* 2016).

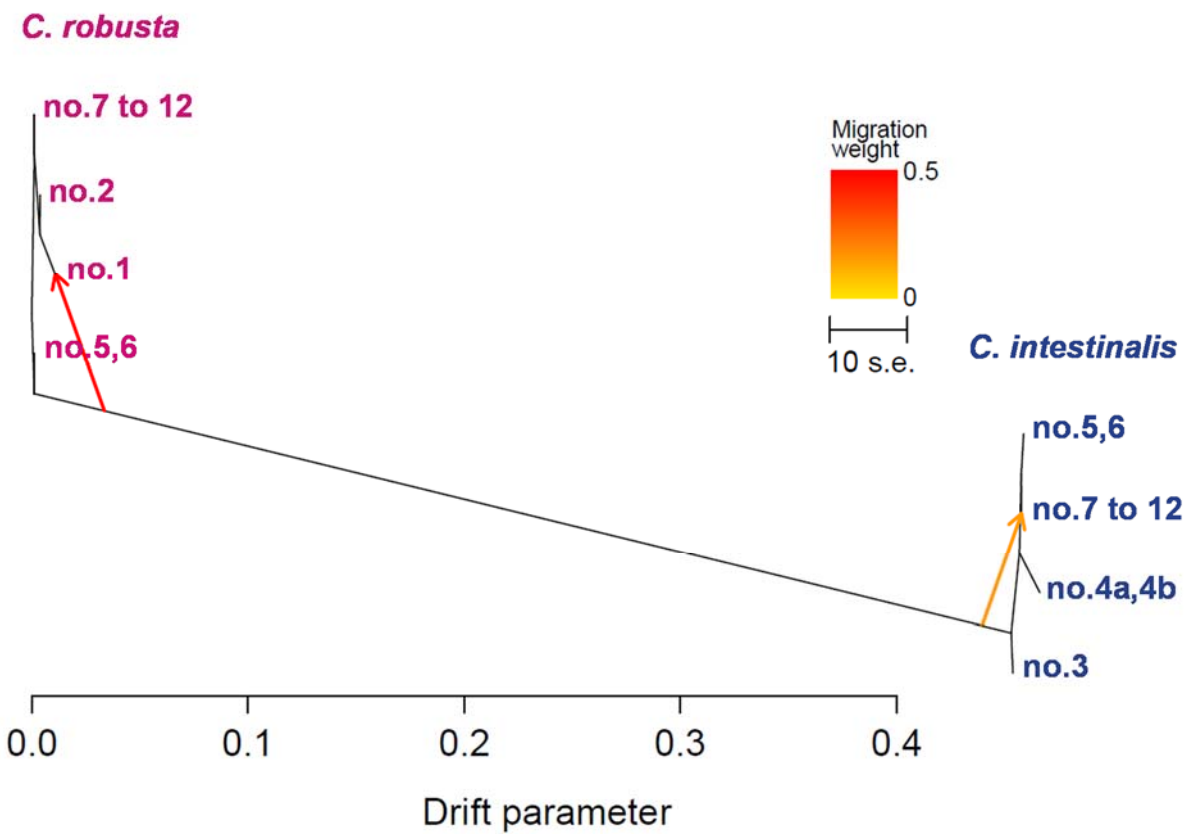


Figure 5. **Population tree inferred by TREEMIX indicating two migration events between *Ciona robusta* and/or *C. intestinalis* populations using the total dataset** (i.e. 310 SNPs). Terminal nodes are labelled by locality number (Table 1). Note that we pooled populations according to regions of sampling (i.e. no.4a and 4b for *C. intestinalis*; no.5 and 6 and no. 7 to 12 for each species) to avoid noises by intra-specific admixture events. Admixture arrows are colored according to the migration weight. The two admixture events significantly improved the model as compared to a situation without migration ($P$<0.001).