

Crossing the Species Barrier: Genomic Hotspots of Introgression between Two Highly Divergent *Ciona intestinalis* Species

Camille Roux,^{*1,2} Georgia Tsagkogeorga,^{1,3} Nicolas Bierne,^{1,2} and Nicolas Galtier¹

¹Institut des Sciences de l'Évolution, Université Montpellier 2, CNRS, Montpellier, France

²Institut des Sciences de l'Évolution, Station Méditerranéenne de l'Environnement Littoral, CNRS, Sète, France

³School of Biological and Chemical Sciences, Queen Mary University of London, London, United Kingdom

***Corresponding author:** E-mail: camille.roux@univ-montp2.fr.

Associate editor: Daniel Falush

Abstract

Inferring a realistic demographic model from genetic data is an important challenge to gain insights into the historical events during the speciation process and to detect molecular signatures of selection along genomes. Recent advances in divergence population genetics have reported that speciation in face of gene flow occurred more frequently than theoretically expected, but the approaches used did not account for genome-wide heterogeneity (GWH) in introgression rates. Here, we investigate the impact of GWH on the inference of divergence with gene flow between two cryptic species of the marine model *Ciona intestinalis* by analyzing polymorphism and divergence patterns in 852 protein-coding sequence loci. These morphologically similar entities are highly diverged molecular-wise, but evidence of hybridization has been reported in both laboratory and field studies. We compare various speciation models and test for GWH under the approximate Bayesian computation framework. Our results demonstrate the presence of significant extents of gene flow resulting from a recent secondary contact after >3 My of divergence in isolation. The inferred rates of introgression are relatively low, highly variable across loci and mostly unidirectional, which is consistent with the idea that numerous genetic incompatibilities have accumulated over time throughout the genomes of these highly diverged species. A genomic map of the level of gene flow identified two hotspots of introgression, that is, large genome regions of unidirectional introgression. This study clarifies the history and degree of isolation of two cryptic and partially sympatric model species and provides a methodological framework to investigate GWH at various stages of speciation process.

Key words: speciation, adaptive introgression, approximate Bayesian computation, *Ciona intestinalis*, invasive species.

Introduction

The study of speciation is a pivotal topic in evolutionary biology, prompted by the question of how time-continuous evolutionary processes, such as genetic drift and natural selection, can lead to discrete genotype and phenotype units, that is, species. According to the biological species concept, the speciation process results in the development of completely isolated gene pools. Although this definition fits well with distant phylogenetic groups with complete evolution of reproductive isolation, the genomes of more closely related entities can be perceived as sieves allowing some genetic exchange, impacting the genomic variation of genealogical relationships even observed for reproductively isolated species (Patterson 2006; Hobolth et al. 2007). This raises questions as to the identification of diverging gene pools, the reconstruction of their demographic history, and the quantification of past and current gene flow rates.

Such questions have stimulated the emergence of divergence population genetics (Kliman et al. 2000), which aims to disentangle the effects of gene flow from those of ancestral polymorphism on the basis of the amount of polymorphism shared by closely related species, and ultimately drawing a historical picture of their relationship. Major methodological developments have been under way to deal with this problem using summary statistic analysis (Wakeley and Hey 1997),

Monte-Carlo Markov chain methods (Nielsen and Wakeley 2001; Hey and Nielsen 2004, 2007; Becquet and Przeworski 2007) and, more recently, approximate Bayesian computation (ABC) methods (Pritchard et al. 1999; Beaumont et al. 2002; Fagundes et al. 2007; Csilléry et al. 2010). Because of their flexibility, ABC methods have successfully proved their potential for studying speciation through comparisons of temporal patterns of introgression (Ross-Ibarra et al. 2009), different scenarios leading to polyploidization (St Onge et al. 2012) and effective population size changes over time (Sjödén et al. 2012; Gattepaille et al. 2013)

These methods have greatly improved our vision of the different forces shaping genome diversification during speciation (Hey and Pinho 2012). It is nevertheless worth noting that most of them are based on models that assume a single common effective migration rate for all loci. However, it is very likely that the gene flow rate varies across loci (Harrison 1993), in a way that depends on the stage of the speciation process (Via 2009). This is because, as gene pools diverge in time, Dobzhansky-Muller incompatibilities are expected to randomly accumulate and create local barriers to gene flow (Wu 2001). Therefore, introgression rates are supposed to be homogeneously high during early stages of speciation, homogeneously zero when the speciation process is completed, and heterogeneous in between, when some but not

all regions of the genome are affected by reproductive isolation factors (Barton and Bengtsson 1986). The genome-wide heterogeneity (GWH) of introgression rates is therefore informative with respect to the position of any pair of populations or species along a continuum of divergence (Via 2009). Orr (1995) predicted that the rate of accumulation of incompatibilities through genetic drift is faster than linear in time, a prediction recently confirmed by experimental counts of genes related to hybrid incompatibilities in fruitflies (Matute et al. 2010) and tomatoes (Moyle and Nakazato 2010).

Recent empirical studies of GWH have mainly focused on the first stages of population divergence using F_{ST} genome scanning. This approach aims at catching speciation in flagrente delicto by identifying “genomic islands of speciation” (Turner et al. 2005; Butlin 2010), that is, loci showing a higher level of differentiation than the average genomic region (Lawniczak et al. 2010). Divergent ecotypes have also been studied to investigate quantitative loci associated with reproductive isolation between closely related species (Presgraves 2003; Tao et al. 2003; Masly and Presgraves 2007; Bikard et al. 2009). At the other end of the continuum of divergence, the study of GWH between highly differentiated gene pools, or even between species, is also of major interest. First, assessment of the level of gene flow between taxa is of major importance for species delineation and taxonomy. Second, gene flow between highly differentiated gene pools is likely to be driven by natural selection, either directional or stabilizing, which can promote local introgression even though most of the genome is unable to cross the species boundary. An analysis of GWH between divergent entities would thus likely reveal interesting adaptive processes (Bierne et al. 2003; Llopart et al. 2005; Castric et al. 2008; Kim et al. 2008; Boon et al. 2009; Whitney et al. 2010; Song et al. 2011; *Heliconius* Genome Consortium 2012; Pardo-Diaz et al. 2012; Gosset and Bierne 2013). Thus, related species can exchange genes through hybrids either because these genes are exceptionally unlinked to barriers to gene flow in contrast to the rest of genome, or because they are part of genomic regions involved in adaptive processes as recently shown in *Heliconius* with colour pattern loci (*Heliconius* Genome Consortium 2012; Pardo-Diaz et al. 2012). Finally, from a fundamental viewpoint, it would be useful to identify the level of sequence divergence above which introgression is definitively impossible, and to compare this threshold across various groups of organisms.

We propose to address these issues by investigating GWH between two highly divergent species for which instances of natural hybridization have been reported, namely the sea squirt (*Ciona intestinalis*). *C. intestinalis* is a self-incompatible species that serves as model system in developmental and evolutionary genomics (Holland and Gibson-Brown 2003; Bourtat et al. 2006; Delsuc et al. 2006; Dunn and Gregorio 2009). It is the first eukaryotic marine species whose genome was completely sequenced (Dehal et al. 2002) after *Takifugu rubripes* (Aparicio et al. 2002). Paradoxically, the study of species delineation and speciation in *Ciona* started recently, when two partially reproductively isolated types, called

C. intestinalis A and *C. intestinalis* B, were identified (Suzuki et al. 2005). A wide spectrum of incompatibilities has been reported between these two cryptic species, depending on the geographic origins of the crossed parents (Caputi et al. 2007). This advanced reproductive isolation likely results from an ancient speciation event, as shown by the high amount of mitochondrial divergence observed between *C. intestinalis* A and *C. intestinalis* B, whose common ancestor probably lived during the Pliocene, 5.3 to 2.6 Ma (Nydam and Harrison 2010). Despite elevated levels of molecular divergence and strong barriers to hybridization in the laboratory, natural cases of introgression between *C. intestinalis* A and *C. intestinalis* B have been reported in sympatric populations along the French Atlantic coast and in the English Channel (Nydam and Harrison 2011). The origin of these mixed *C. intestinalis* populations is still being debated. It was recently suggested that the actual worldwide distribution of *C. intestinalis* is the result of human activities and worldwide boat travel (Ramsay et al. 2008; Zhan et al. 2010). These hypotheses should be specifically tested, especially as the global spread of *Ciona* populations has become an ecological and economic concern (Robinson et al. 2005; Daigle and Herbingier 2009).

Here, we took advantage of the robustness and flexibility of ABC methods (Beaumont et al. 2010) to quantify GWH in *C. intestinalis* using next-generation sequencing technology. To investigate the historical genomic relationship between the two diverging species, we analyzed patterns of polymorphism among species and populations based on individual transcriptome sequencing. This data set was analyzed with simulation-based approaches to estimate the following: 1) the best-fitting demographic model of speciation of the genomic background, 2) the joint demographic parameters shared by all loci, 3) the shape and intensity of GWH, and 4) the genomic localization of past gene flow events. Our study revealed gene flow between *C. intestinalis* A and *C. intestinalis* B in recent times, while highlighting the need to take GWH into account in ABC analyses, as its omission can lead to false conclusions about former speciation processes between diverging entities.

Results

Patterns of Polymorphism and Divergence in *C. intestinalis*

To investigate GWH, we obtained a polymorphism data set from individual transcriptomes resequenced in 20 sea squirt individuals, that is, 10 *C. intestinalis* A individuals from the Mediterranean Sea and Pacific Ocean and 10 *C. intestinalis* B individuals from the North Sea, the western and eastern Atlantic coasts of France and the Canadian Maritimes. We analyzed 852 loci for which polymorphism data were available in the two species, with a total alignment length of 270 kb. Levels of synonymous diversity, measured by either Watterson's θ_W (Watterson 1975) (fig. 1A) or π (Tajima 1983; supplementary table S1, Supplementary Material online), were significantly different between the two *Ciona* species (Wilcoxon signed-rank test, $V=9,386.5$, P value < 0.0001 for θ_W), with *C. intestinalis* B (average = 0.0539,

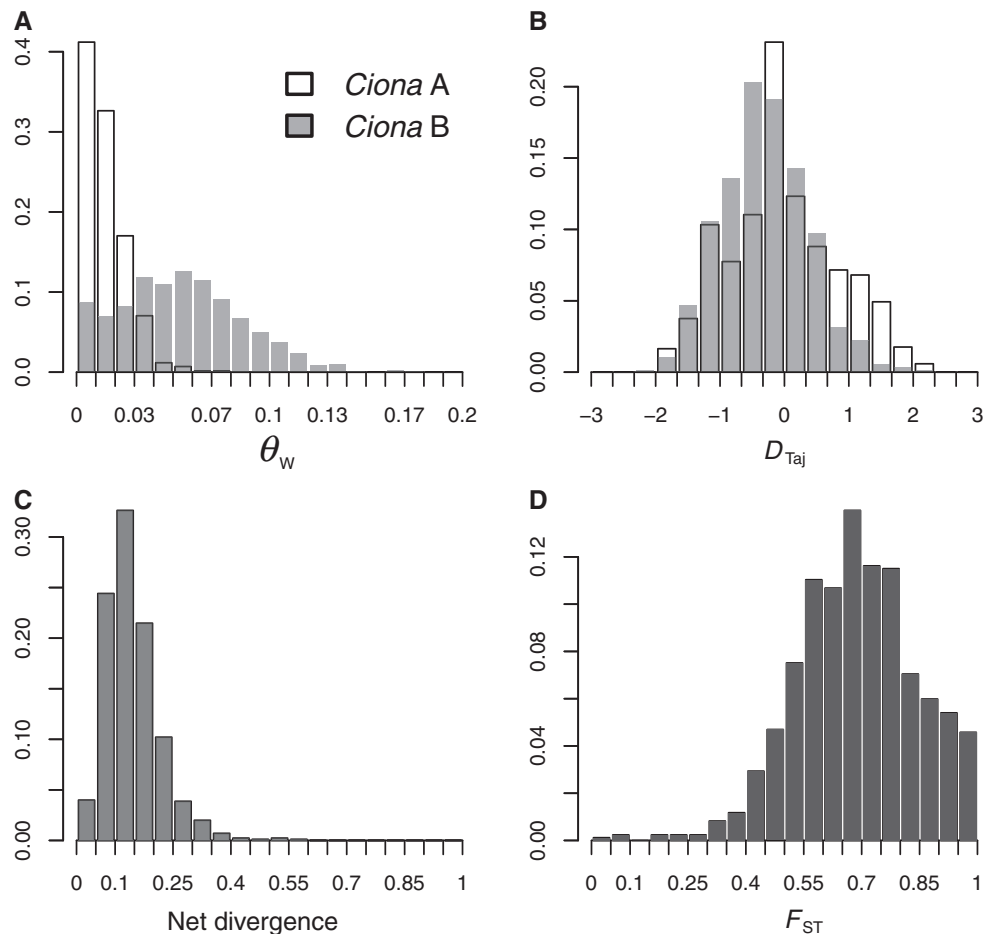


Fig. 1. Patterns of polymorphism and divergence between *Ciona* species. Histograms show the empirical distributions of (A) Watterson's θ diversity estimator. (B) The site frequency spectrum measured by Tajima's D . (C) The net inter-specific divergence. (D) Between-species differentiation measured by F_{ST} . Each statistics are measured at synonymous positions.

standard deviation [SD] = 0.0313 for θ_w) being more polymorphic than *C. intestinalis* A (average = 0.0143, SD = 0.0108 for θ_w). To test whether this difference in nucleotide diversity was due to a recent demographic expansion affecting the whole range of *C. intestinalis* B or a severe bottleneck specific to *C. intestinalis* A, we calculated Tajima's D for each locus and compared the observed values with neutral expectations obtained from 1,000 coalescent simulations conditioned on the locus-specific number of segregating sites. Tajima's D multilocus (Ramos-Onsins et al. 2004) showed no departure from the standard neutral model (SNM) for both *C. intestinalis* A (average = -0.01939 , $P = 0.4902$) and *C. intestinalis* B (average = -0.3136 , $P = 0.5977$), even though Tajima's D was significantly more negative in *C. intestinalis* B than in *C. intestinalis* A (fig. 1B; Wilcoxon signed-rank test, $V = 229,747$, P value < 0.0001). In addition, not a single locus exhibited a significant deviation from the SNM when measured by Tajima's D , Fu and Li's F^* and D^* , and only R2 (Ramos-Onsins and Rozas 2002) revealed a departure from SNM for 10 loci (supplementary table S2, Supplementary Material online). These results are consistent with the hypothesis of demographic equilibrium in the two species, which is the ideal situation for studying natural selection. However, we note that this result may in part reflect the lack of power of neutrality tests in case of

insufficient sampling of structured populations, as suggested by simulations. Stadler et al. (2009) demonstrated that a genome-wide departure from SNM, for instance in the case of population expansion, may be misdetected when neutrality tests are performed on pooled individuals sampled from structured populations. These effects of sampling are highly sensitive to the age of populations as well as the growth rate, that is, intraspecific parameters that require larger samples to be investigated. The number of substitutions per synonymous site between these species (fig. 1C; average = 0.1444, SD = 0.0721) was very close to previous estimates reported for mitochondrial DNA (Nydam and Harrison 2010), and thus confirming the strong species differentiation measured by F_{ST} (fig. 1D; average = 0.6870, SD = 0.1555). Although some trans-specific polymorphism segregated at 282 loci out of 852 (fig. 2), the two species were well separated by the first factorial axis, even in the European populations that are now in contact in the English Channel (supplementary fig. S1, Supplementary Material online).

Strong Statistical Support for GWH during a Secondary Contact

The shared polymorphism that was noted could be explained by incomplete lineage sorting (Clark 1997), secondary

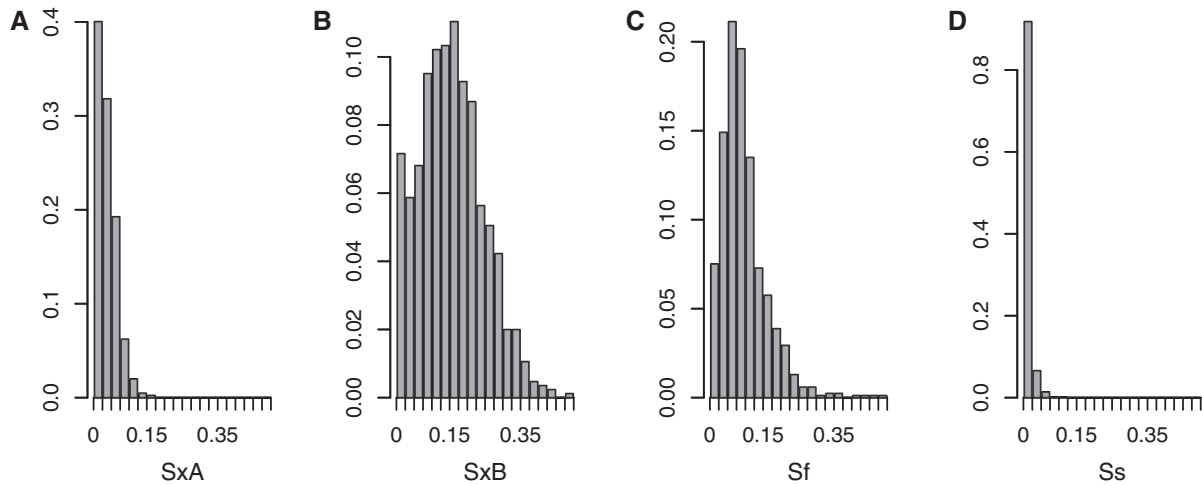


Fig. 2. Distribution of variable positions in *Ciona* comparisons. Distribution of the proportion of synonymous sites (A) exclusively polymorphic in *C. intestinalis* A; (B) exclusively polymorphic in *C. intestinalis* B; (C) differently fixed between the two species; (D) with a biallelism shared by both species.

introgression, or both, with introgression being either homogeneous or heterogeneous across loci. We used a model-selection approach to identify a relevant speciation model in an ABC framework. We performed coalescent simulations under various models to compute the expected distributions of summary statistics related to polymorphism and divergence patterns (fig. 3). Then, we compared the simulated statistics with the observed ones calculated from the data.

We considered three models involving migration: constant migration (CM), ancient migration (AM), and secondary contact (SC). Within each model, GWH was tested by the comparison of two alternative introgression versions assuming either a single introgression rate shared by all loci (Homo), or independently sampled introgression rates across loci (Hetero). In all three cases, the Hetero version clearly fitted our data set better than the Homo version, as supported by the high relative probability that Hetero was the correct version given our observed posterior probabilities (table 1). We conducted a simulation analysis to assess the robustness and power of this approach (supplementary fig. S2, Supplementary Material online). One thousand pseudo-observed data sets were simulated under each of the six earlier-described migration models and subjected to our ABC-based Homo vs. Hetero analysis. Similar to the analysis of the observed data set, we allowed the strength of GWH measured by the variance of the Beta distribution to range from 2.5×10^{-5} to 0.25 among the Hetero versions. For the two models that allowed ongoing migration (CM and SC), the percentage of correct assignment was 100% for both the Homo and Hetero alternative versions. When migration events were assumed to be restricted to the early speciation stages (AM), our approach correctly supported the Homo version in 92.8% of the 1,000 pseudo-observed Homo data sets, and the Hetero version in 87% of the 1,000 pseudo-observed Hetero data sets (supplementary fig. S2, Supplementary Material online). ABC therefore appears to be a robust approach to gain insight into the degree of variation in introgression rates between two genomes.

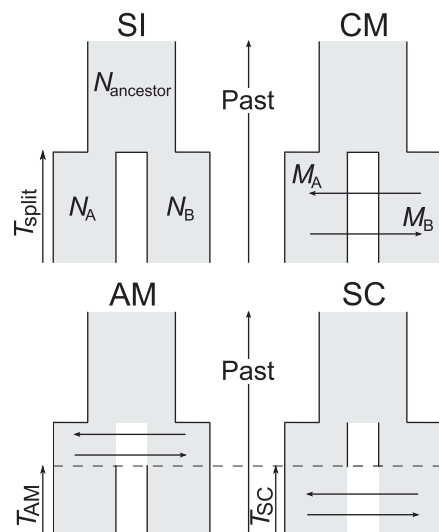


Fig. 3. Alternative scenarios of speciation for *C. intestinalis* A and B. Four classes of models with different temporal patterns of migration have been compared: SI, CM, AM, and SC. Four parameters are shared by all models: T_{split} is the number of generations since the speciation time; N_{ancest} , N_A , and N_B are the number of effective individuals in the ancestral population, *Ciona intestinalis* A and *C. intestinalis* B respectively. T_{AM} is the number of generations since the two nascent species have stop to exchange migrants. T_{SC} is the number of generations since the two daughter species experiment a SC after a period of isolation. The effective migration rates M_A and M_B are expressed in $4Nm$ units, where m is the proportion of a population made up of migrants from the other population per generation.

We then tested the various hypotheses regarding the history of gene flow between *C. intestinalis* A and *C. intestinalis* B by independent comparisons of the Hetero versions of AM, CM, and SC models against the strict isolation (SI) model (table 2). Although the posterior probability of SI (0.7534) was greater than AM (0.2466), the two CM and SC models with ongoing migration outperformed SI with a posterior probability of 0.9998 for each of them. The support for

Table 1. Test for Genomic Variation in Introgression Rates.

	Within Models		Robustness
	Homo	Hetero	
CM	0.0006	0.9994	1
AM	0.0167	0.9833	0.9999
SC	0.0001	0.9999	1

NOTE.—Values represent the posterior probabilities of each alternative models of introgression, with interlocus variation in effective migration rates (hetero) or not (homo). The robustness is the probability for each comparison that the best-supported alternative model is the correct model given the observed posterior probability, computed by ABC analysis of 1,000 pseudo-observed data sets simulated in each alternative model.

Table 2. Classification of Models of Speciation.

	Comparisons with SI		Comparison CM vs. SC	
	P(SI)	Robustness	P	Robustness
AM (Hetero)	0.7534	1	—	—
CM (Hetero)	0.0002	1	0.0016	0.9886
SC (Hetero)	0.0002	1	0.9984	

NOTE.—Values represent the posterior probabilities of the SI model against AM, CM, and SC and the associated robustness. The two best-supported models CM and SC are then compared.

models with ongoing migration was robust to differences in the number of parameters and in the variance of the simulated statistics, as indicated by a probability to reject SI given our observations of $P(\text{SI}) = 0.0002$ equal to one in both comparisons.

We finally investigated the chronology of migration by comparing CM-Hetero with SC-Hetero models (table 2) in the same methodological framework. SC-Hetero better fitted our data set than CM-Hetero, with a highly significant posterior probability of 0.9984, suggesting that the two taxa had experienced a period of complete isolation followed by an SC. To ensure us that our analysis was not biased by a possible linkage between loci, our model-selection procedure was repeated three times using 85 randomly chosen loci ($\approx 10\%$ of the original data set). The SC-Hetero model was supported in all three replicates.

Interestingly, the statistical support for ongoing migration would have been overlooked if GWH had not been considered. When only the Homo versions of our migration models were pairwise compared to the SI model, we observed a strong rejection of CM ($P[\text{SI}] = 0.9999$), and a posterior probability of SC ($P[\text{SC}] = 0.5398$) only slightly superior to that of SI ($P[\text{SI}] = 0.4602$; supplementary table S3, Supplementary Material online). However, comparing the four models in the same analysis instead of conducting pairwise model comparisons clearly supported SI ($P[\text{SI}] = 0.6525$) to the detriment of CM ($P[\text{CM-Homo}] = 0.0011$), AM ($P[\text{AM-Homo}] = 0.0012$), and SC ($P[\text{SC-Homo}] = 0.3452$), whereas SC was still the best-fitted model ($P[\text{SC-Hetero}] = 0.9921$) when simultaneously compared with SI, CM-Hetero, and AM-Hetero. Thus, taking the variance in the introgression rate across loci into account in our best-fit procedure seems essential for a

suitable assessment of the demographic history of these two taxa. The risk of neglecting these effects would lead to contradictory conclusions about the mode of speciation tested through a pair of related species.

Inference of Historical Parameters under the SC-Hetero Model

We examined parameter estimates under the best-fitting SC-Hetero model of *C. intestinalis* A/B divergence. The posterior distributions of parameters built from 2,000 accepted simulations were highly differentiated from their prior distributions, suggesting that the data offer adequate information, and that we appropriately explored the parameter space (fig. 4). In this analysis, the mutation rate was assumed to be constant over time and across loci, and equal to 1×10^{-8} bp/year as recently estimated (Tsagkogeorga et al. 2012). Estimates of the effective population sizes suggest an approximately 3.75-fold larger population in *C. intestinalis* B (883,000, 95% confidence interval [CI]: 748,000–1,022,000) than in *C. intestinalis* A (235,000, 95% CI: 115,000–395,000). Our analysis also suggests that both daughter populations were smaller than the ancestral one (2,000,000, 95% CI: 1,606,000–2,220,000). Regarding the timing, our estimations indicate that the split in the ancestral *C. intestinalis* population probably occurred between 2.7 and 5.5 Ma (point estimate: 3.76 Ma). Subsequent to this ancestral divergence, the two *C. intestinalis* species would have evolved separately over a long period until a recent SC occurred 15,500 years ago (95% CI: 4,300–56,800).

We then investigated the predicted distributions of introgression rates from *C. intestinalis* B into A ($M_A = 4N_A m_A$) and from *C. intestinalis* A into B ($M_B = 4N_B m_B$) throughout the genomes, based on 2×10^6 random sampling from 2,000 accepted simulations assuming free recombination between loci, with GWH being modeled by two joint Beta distributions in our ABC analysis. Both intragenomic distributions were leptokurtic and the inferred medians for M_A (0.079) and M_B (0.05) were in the same order of magnitude (fig. 5). The proportion of loci with an expected effective migration rate under the value of one (considered as a limit to differentiation [Templeton 2006]) was 92% and 90% for M_A and M_B , respectively. The effective gene flow was predicted to vary slightly more across loci into *C. intestinalis* B (95% of the genomic distribution was between 0 and 6.6074) than into *C. intestinalis* A (95% of the genomic distribution was between 0 and 2.4272). We then compared observed and simulated summary statistics under a goodness-of-fit procedure from the joint-posterior distribution, and found that our estimated SC-Hetero model fitted well to the data (supplementary fig. S3, Supplementary Material online) with the exception of two statistics related to Tajima's *D*, which slightly departed from the expectations in *C. intestinalis* B.

Genomic Hotspots of Gene Flow

We finally tested introgression locus by locus by comparing different migration models, that is, a complete isolation model (M_0 with $M_A = M_B = 0$), models allowing migration

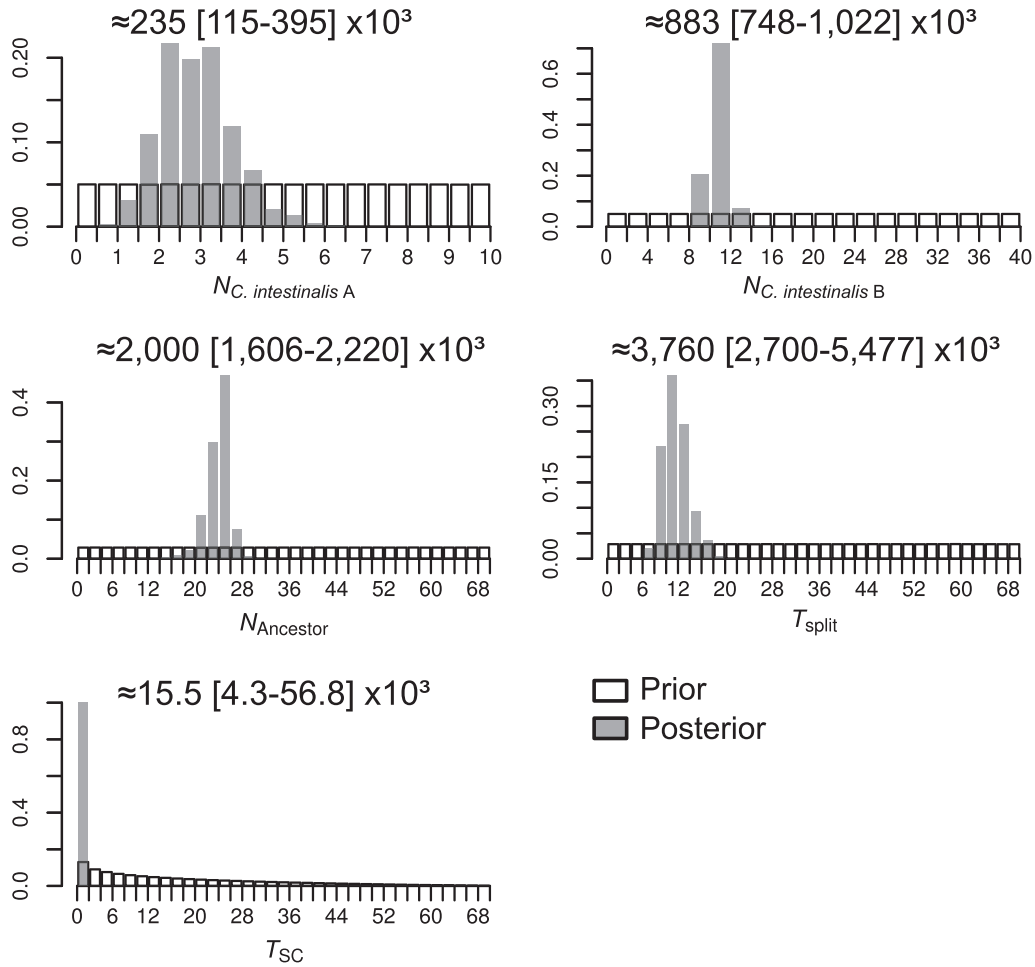


Fig. 4. Parameter estimates of the best supported model of speciation SC. Prior and posterior distributions for parameters shared by all loci are open and shaded, respectively. The median and the 95-highest posterior density of each parameter is given along with its conversion into number of effective individuals or years.

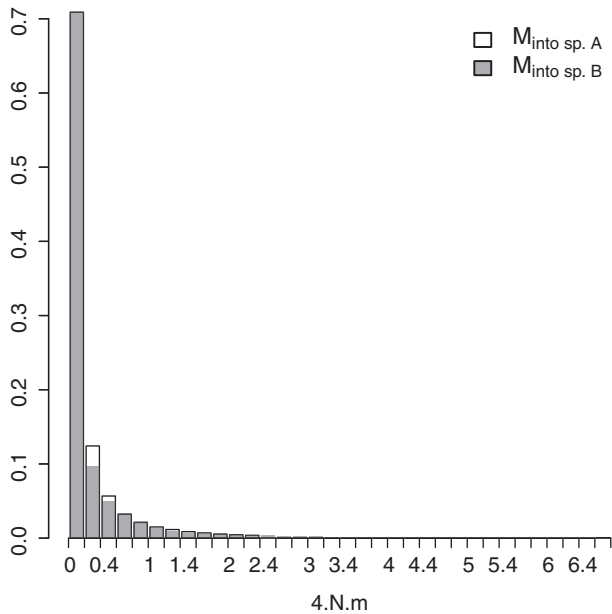


Fig. 5. Estimated genomic distributions of introgression rates into *Ciona intestinalis* A and B. Distributions are obtained after randomly sampled 1,000 values from each 2,000 Beta distributions retained by ABC analysis.

only into *C. intestinalis* A (M_1 , with $M_A > 0$ and $M_B = 0$) or *C. intestinalis* B (M_2 , with $M_A = 0$ and $M_B > 0$) and a model assuming introgression in both directions (M_3 , $M_A > 0$ and $M_B > 0$). For each sequenced locus, we simulated the four alternative introgression models 1.5×10^6 times and applied our model choice procedure to characterize its migration history (fig. 6; supplementary table S4, Supplementary Material online). A preliminary analysis using simulated pseudo-observed loci revealed that this approach was straightforward for detecting the presence of introgression as well as its absence, although it was slightly more difficult to distinguish between the three models with ongoing migration M_1 , M_2 , and M_3 (supplementary fig. S4, Supplementary Material online).

In line with the multilocus estimated introgression rate distribution, the locus-specific analysis pointed to the existence of a strong barrier to gene flow, with a majority of loci ($\approx 79\%$) being associated with the M_0 model. Nonetheless, we found slight differences with the multilocus analysis concerning the introgression symmetry. The proportion of loci supporting the M_1 model ($\approx 10.5\%$) was similar to the proportion of loci best supported by the M_2 model ($\approx 10\%$), and only a small number of loci supported the

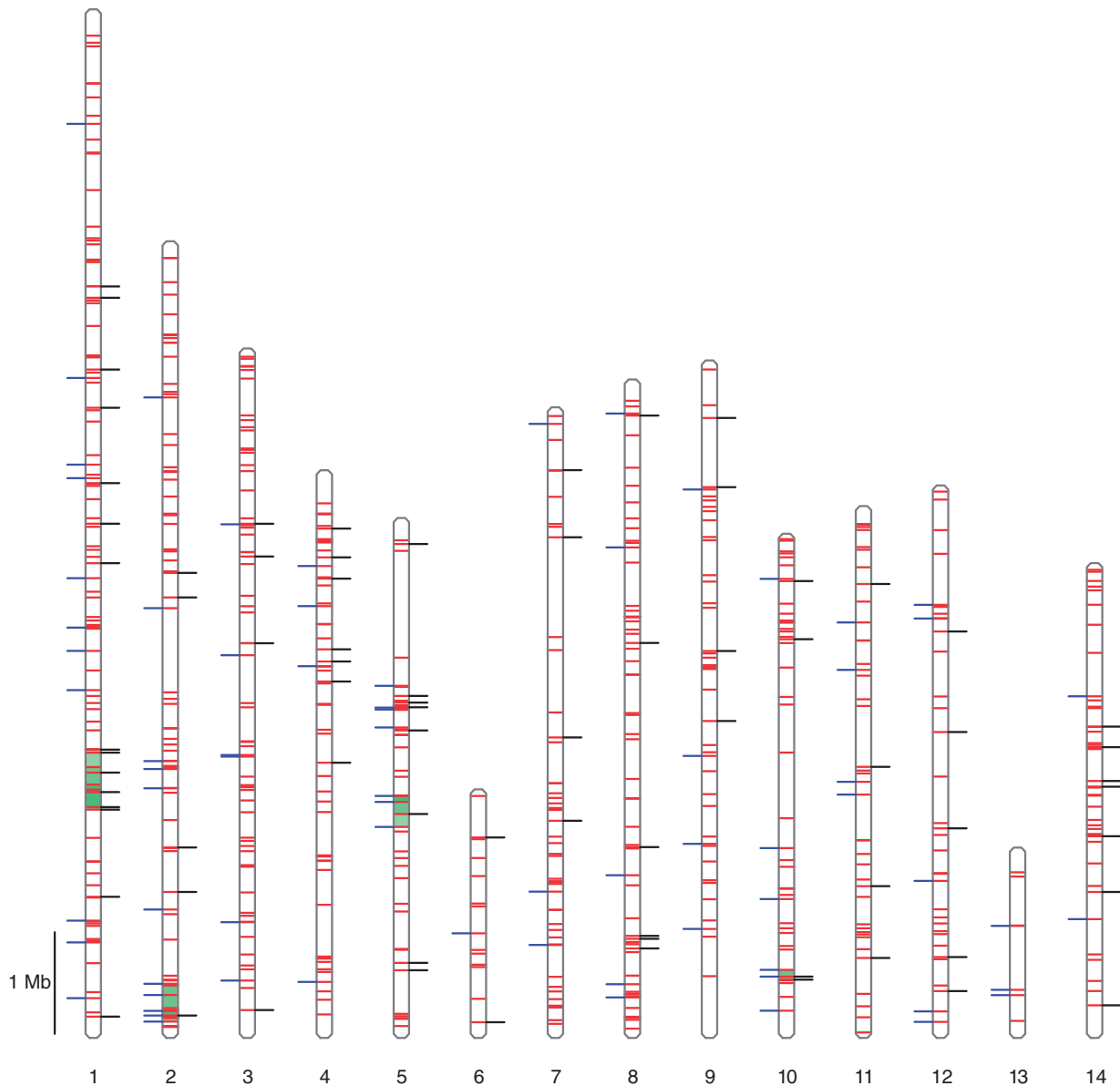


Fig. 6. Physical mapping of historical migration events in *Ciona intestinalis* genome. Resequenced loci are positioned by red segments along the chromosomes. Introgression events into *C. intestinalis* A are indicated by the blue segments. Introgression events into *C. intestinalis* B are indicated by the black segments. Regions revealed as significantly enriched in migrant loci by the sliding-window analysis are indicated by green boxes.

M3 model ($\approx 0.5\%$). All loci showing no fixed differences were associated with models M1, M2, or M3, and 97% of the 570 loci showing no shared polymorphism were associated with model M0, suggesting that the observed shared polymorphism resulted from genetic exchanges between the two *C. intestinalis* species, not from the retention of ancestral polymorphism.

Finally, the existence of genomic hotspots of introgression was investigated by building a genomic map from the physical genome of the analysed loci for regions of reduced prevalence of the M0 model using a sliding-window strategy and permutation tests. Each overlapping window contained five successive contigs. We excluded windows containing two contigs more distant than 150 kb. Four genomic regions enriched in introgressed loci were thus detected on chromosomes 1, 2, 5, and 10. The hotspot on chromosome 1 contained six

sequenced loci supported by model M2, thus indicating introgression from *C. intestinalis* A into *C. intestinalis* B. The one on chromosome 2 contained four sequenced loci supported by M1, plus one symmetric locus. We called “genomic hotspots of introgression” as the regions enriched in loci crossing the species barrier in one direction. The two other regions include a mixture of loci introgressing both ways, containing four and three adjacent loci on chromosomes 5 and 10, respectively. We thus inferred that $\approx 11\%$ of the loci that show evidence of migration is located within the four detected hotspots of introgression. The list of genes located within these detected hotspots of introgression is given as [supplementary material \(supplementary table S5, Supplementary Material online\)](#) as well as their functional annotation (propagated from human orthologs, when available).

Discussion

The prevalence of gene flow between differentiated species has been fully recognized as an evolutionary force contributing to the diversity of plants (Soltis and Soltis 2009), vertebrates (Brumfield et al. 2001; Helbig et al. 2001; Won and Hey 2005; Carling and Brumfield 2008), invertebrates (Cianchi et al. 2003; Mullen et al. 2008), and fungal systems (Gladieux et al. 2008, 2011), but its role in evolution is still matter of debate (Arnold 2004; Barton 2001, 2013; Abbott et al. 2013). The fate of an allele introgressing into a new gene pool is known to be highly dependent on the nature and strength of natural selection acting on its genomic neighbourhood. The indirect effects of selection at linked loci presumably explain the variation in introgression rate among genomic regions—from strong barriers to gene flow in the neighbourhood of barrier loci (so-called genomic islands of differentiation) to basal gene flow in regions devoid of barriers and maximal gene flow at, and around, adaptively introgressed loci. In demographic studies, genomic heterogeneity in introgression rates is typically seen as being a confounding factor, not a relevant biological parameter. Here, by considering a distribution of the indirect effects of natural selection on the effective migration rate, across *C. intestinalis* A and B genomes when analyzing polymorphism and divergence patterns, we gained insight into numerous aspects of the demographic history of these taxa.

Population Genomics of *C. intestinalis* Divergence

The observed molecular data are best explained by a speciation event which occurred ≈ 3.8 Ma, confirming the suggestion made by Nydam and Harrison (2007), based on a molecular clock approach, of a common ancestor to *C. intestinalis* A and B living during the Pliocene (from 5.332 to 2.588 Ma). Knowlton et al. (1993) and Knowlton and Weigt (1998) proposed that a pulse of speciation events occurred in shallow-water species during this period, which corresponds to the time of complete closure of the connection between the Western-Atlantic and Eastern-Pacific oceans by the Isthmus of Panama, 3–3.5 Ma. The geographical isolation of species on each side of the isthmus was associated with reproductive isolation in several shallow-water species of temperate oceans (Lessios 2008), even though this geological event alone does not appear to explain all of the reported cases of speciation in that geographic area (Hurt et al. 2009). Thus, we speculate that the closure of the Isthmus of Panama might be the cause of the initial split between the Pacific (A) and Atlantic (B) lineages of *C. intestinalis*, even though we obviously have no direct proof for this scenario.

Our ABC-based comparison of models strongly supports the existence of a recent contact between the two *C. intestinalis* species, after a long divergence period in complete isolation. It is worth stressing that this recent gene flow episode is overlooked by models assuming a constant effective migration rate across loci, the Homo versions of our demographic models are equivocal regarding the existence of an SC between the two species. We draw the following genomic picture of ongoing introgression: 1) a minority of loci

($\approx 20\%$) crossed the species barrier, as shown by locus-specific model comparisons; 2) almost all of the introgressing loci ($\approx 99\%$) experienced unidirectional allele exchange, with a similar number of loci introgressing from species B into species A than from species A into species B; 3) introgression rates are very low, with a median of 0.079 from B into A, and 0.05 from A into B.

According to our Hetero-SC model, the SC would have started approximately 15,000 years before present. At this time, the world climate was warming after the last glacial maximum, which resulted in the retreat of ice sheets and major changes in the circulation of oceanic currents (Hu et al. 2010), presumably impacting the demographic history and geographic range of species (Hewitt 2000). Regardless of the exact cause of the SC, our model suggests that natural hybridization between *C. intestinalis* A and B occurred before human activities began to substantially perturb the range and dispersal of the various *C. intestinalis* entities (Lambert and Lambert 1998; Ramsay et al. 2008). Importantly, the principal component analysis indicated that the three *C. intestinalis* B populations that were sampled are genetically equally distant from *C. intestinalis* A, irrespective of their geographic proximity to current *C. intestinalis* A populations. This result is consistent with the hypothesis that the gene flow from A to B that we detected occurred long enough in the past to spread across the whole range of *C. intestinalis* B.

It seems surprising that the two gene pools have remained in complete isolation for > 3 My, as supported by our analysis, even though the potential for hybridization and genetic exchange has been preserved until present day. This scenario seems to require a situation of uninterrupted allopatry between the two entities, suggesting that the geographic distribution of the two species during the Pleistocene differed from the current one. Note also that species A and B might not be the only players in the *C. intestinalis* system. Two divergent genetic lineages were recently identified in the Mediterranean Sea (Zhan et al. 2010). These taxa, and possibly additional undescribed or extinct ones (as there are almost no fossil records on these soft-bodied animals), may have occupied larger oceanic areas over the last 3 My, and we did not evaluate the possible effects of gene flow from these additional *Ciona* lineages on our demographic inferences.

We acknowledge, finally, that the set of models explored in this study does not cover the entire range of potential divergence scenarios, a general caveat of divergence population genetics. One could imagine, for instance, that *C. intestinalis* A and B species experienced a series of brief episodes of contact, at various time points during their divergence. Our SC model provided a reasonable fit to the data and is strongly preferred as compared with complete isolation, continuous migration, and AM, but it clearly does not provide a definitive picture of the history of the taxa considered.

Crossing the Species Barrier

The two species analyzed in this study showed an average of 14.4% synonymous sequence divergence. This is equivalent to the divergence between humans and marmosets

(Perelman et al. 2011). Although detecting significant amounts of gene flow between so differentiated gene pools was unexpected, previous empirical and field studies have nevertheless shown that inter-specific crosses yield some viable F1 offsprings (Suzuki et al. 2005; Caputi et al. 2007; Sato et al. 2012). In the light of these observations, the genome-wide barrier to gene flow we infer may seem paradoxical. However, we suggest that the two results 1) do not concern the same time scale and 2) are consistent with the dominance theory of postzygotic isolation (Turelli and Orr 1995, 2000). If the barriers to gene flow were mostly due to recessive Dobzhansky–Muller incompatibilities, then viable hybrids are expected to occur after one generation of crossing, whereas the expression of recessive mutations in subsequent generations of hybridization (F2, backcrosses) should prevent the long-term establishment of gene flow and introgression between the two gene pools. The investigation of the genetic basis of postzygotic isolation between two species of marine mussels, other ocean-dwelling broadcast spawners, with a shorter period of isolation (1–2 Ma), have corroborated the existence of a high number of recessive Dobzhansky–Muller substitutions scattered in the genome and able to impose a strong and genome-wide barrier to gene flow (Bierne et al. 2006). *C. intestinalis* therefore proves to be an ideal model for studying the introgression process in the terminal stage of speciation and to determine the extent to which these events are favored by natural selection. The inferred introgression rate distribution between the two *C. intestinalis* genomes (both ways) was highly leptokurtic, as expected in the final stages of evolutionary divergence, after the build-up of numerous barriers to gene flow throughout the genome. Our analysis suggests that introgression concerned isolated scattered loci, as well as two unidirectional hotspots, characterized by an unexpectedly high number of loci exchanging alleles in one way.

This raises the question of the evolutionary forces driving introgression in these different genomic regions. Under a pure Dobzhansky–Muller model of incompatibility accumulation, genomic hotspots of introgression might result from a depletion of barrier loci and/or a high rate of recombination locally on the chromosome. Giant genomic islands of differentiation (or “continents,” Michel et al. 2010) would affect most of the genome while introgression would proceed only in the few remaining genome regions devoid of any barrier loci. Under this hypothesis, however, it is unclear whether introgression should be symmetrical or if it could be unidirectional as we observed. Alternatively, a number of recent studies have shown that natural selection can enhance introgression in large genomic regions, estimated to be around ≈ 60 kb in length in *Arabidopsis* (Castric et al. 2008; Goubet et al. 2012), ≈ 100 kb in *Heliconius* (*Heliconius* Genome Consortium 2012; Pardo-Diaz et al. 2012), and ≈ 5 Mb in *Neurospora* (Sun et al. 2012). High resolution data sets focusing on the detected hotspots would be required to discriminate between these two hypotheses in *C. intestinalis* and elucidate the evolutionary forces at work (Bierne 2010). Determining the extent and cause of introgression is a major evolutionary biology issue as it is still not understood

whether adaptive introgression is a spectacular illustration of adaptation in a structured population but of little evolutionary relevance (Barton 2013), or whether it could really be a source of evolutionary novelties and big adaptive jumps (Mallet 2007; Abbott et al. 2013).

To gain insight into the impact of natural selection on introgression, a candidate gene approach could be adopted by focusing on genomic regions linked to loci undergoing adaptation (Whitney et al. 2006; Boon et al. 2009) or negative frequency-dependent selection (Schierup et al. 2000), and comparing the estimated introgression rates with the genomic average, or the genomic distribution. In *Ciona*, the two loci responsible for self-incompatibility (Harada et al. 2008) are good candidates for a locally elevated introgression rate—an allele that would enter a genetic pool in which it was so far previously absent would be strongly favored by negative frequency-dependent selection. Roux et al. (2012), however, suggest that even moderate recombination rates are sufficient to break the linkage between a self-incompatibility locus and neighboring regions. Assessing the influence of natural selection on introgression in *C. intestinalis* self-incompatibility loci would therefore require sequencing of DNA fragments in close proximity to (i.e., a few kilobases away from) the targeted S-loci. In the data set analyzed here, the locus closest to a self-incompatibility locus was located 11.5 kb away from the S-locus on chromosome 7 while showing no specific introgression pattern.

Finally, one might ask whether the existence of interspecific gene flow contributes to the very high morphological similarity between *C. intestinalis* A and B, and more generally between the many cryptic species that have recently been discovered (Bickford et al. 2007). Stabilizing selection and/or convergent evolution of morphological traits could possibly explain why two distant populations do not exhibit any detectable morphological differences. We noted, however, that *C. intestinalis* A and B had presumably evolved in spatio-temporally variable environments over the last 3–4 My. Alternatively, the lack of morphological divergence between the A and B species might have resulted, at least in part, from the introgression of a large number of unlinked loci, each with relatively minor effects on morphology (Liu et al. 2012).

In summary, we show that taking into account the realistic genomic heterogeneity of introgression rates is of prime importance in demographic analysis. The risk of overlooking variations in indirect effects of natural selection on loci is that an inappropriate biological speciation model may be supported, ultimately biasing inferences of selective pressures in the species under study. We also showed that two species can produce viable hybrids despite a genomic fraction that introgressed in one direction approximating only 10%. Similar investigations on other biological models will allow us to quantify how common is this observation. Our data also allowed us to analyze an important aspect of speciation, the shape of GWH in an extreme speciation stage between two highly divergent species. Future investigations could also be conducted on numerous pairs of taxa showing variable levels of divergence to ultimately highlight the speciation process with respect to the relationship between molecular

divergence and the strength of barriers to gene flow (Feder et al. 2012),

Material and Methods

Sampling and Sequencing

For *C. intestinalis* A, we sampled 10 adult individuals from three natural populations (supplementary table S6, Supplementary Material online), two European, France (Etang de Thau: $N = 4$), and Spain (Barcelona: $N = 3$), and one population from the west coast of United States in California (Santa Barbara: $N = 3$). For *C. intestinalis* B, we used the RNA-Seq data previously generated by (Tsagkogeorga et al. 2012), including 10 illumina-sequenced transcriptomes from individuals sampled in Canada (Nova Scotia: $N = 3$) and Norway (Bergen: $N = 5$). We also included individuals from the current sympatric zone in France (Concarneau: $N = 2$). Total RNA extractions were performed from pooled fresh tissues of gonads and muscles for each individual, using the RNeasyPlus Kit (Qiagen, Chatsworth, CA) and purification over spin columns following (Gayral et al. 2011). GATC Biotech conducted the cDNA library preparation, the tagging and sequencing of all samples on a Genome Analyzer II, according to standard Illumina protocols (Illumina, Inc.).

Mapping and Data Set Assembly

The in silico approach described in (Tsagkogeorga et al. 2012) was followed for annotating the RNA-Seq data for *C. intestinalis* A. In brief, we characterized the 10 novel transcriptomes by direct read mapping to the reference transcriptome of *C. intestinalis*, downloaded from Ensembl (<http://www.ensembl.org/>, last accessed April 23, 2013), using Burrows–Wheeler alignment tool (Li and Durbin 2009). The identified transcripts were next filtered based on the average sequencing depth and their expression across individuals. Coding sequence alignments were finally built upon a subset of 1,602 previously assembled orthologous protein-coding gene data sets for *C. intestinalis* A and B (Tsagkogeorga et al. 2012). Here, loci that were expressed in at least four individuals *C. intestinalis* A and four *C. intestinalis* B at a minimum $10\times$ coverage were selected for further analysis.

Data Analysis

We restricted our analysis to contigs successfully assembled in at least four diploid individuals belonging to both *Ciona* species and longer than 35 codons after exclusion of codons containing indels and exclusion of positions containing more than two segregating alleles. For each locus, we computed an array of statistics related to polymorphism and divergence at synonymous sites: nucleotide π (Tajima 1983), Watterson's θ_W (Watterson 1975), F_{ST} (estimated as $1 - \pi_S/\pi_T$, where π_S is the mean pairwise diversity for both species and π_T is the pairwise diversity calculated from the total alignment), Tajima's D (Tajima 1989), Fu and Li's D^* and F^* (Fu and Li 1993), R_2 (Ramos-Onsins and Rozas 2002), total interspecific divergence, net interspecific divergence, the number of biallelic positions exclusively polymorphic in *C. intestinalis* A (S_{xA}), in *C. intestinalis* B (S_{xB}) or shared by both species (S_s), and the number of fixed differences

between *C. intestinalis* A and *C. intestinalis* B. This set of summary statistics does not rely on genotype phasing, and has been widely used in the past for demographic inference (Wakeley and Hey 1997; Becquet and Przeworski 2007; Fagundes et al. 2007; Ross-Ibarra et al. 2008; Roux et al. 2011). Locus-specific tests of neutrality were performed by contrasting the observed values to neutral distributions obtained after 1,000 coalescent simulations conditioned on the locus-specific number of segregating sites (Ramos-Onsins et al. 2004; Ramírez-Soriano et al. 2008). Neutral samples were simulated under a Wright–Fisher demographic model, assuming a panmictic population of constant size.

ABC Analysis

Coalescent Simulation

To investigate the demographic history between *C. intestinalis* A and *C. intestinalis* B, we compared four speciation models using an ABC framework (Tavaré et al. 1997; Beaumont et al. 2002), all the tools used to simulate models as well as the command lines are available on the website <http://www.abcgwh.siteweb.ch> (last accessed April 23, 2013). Among these models, three assume a nonzero rate of allele exchange through migration events at different timescales (continuous migration, AM and SC; fig. 3), the fourth model is the standard SI model. All models consider an instantaneous split of the ancestral population of effective size N_{anc} into two daughter populations of independent constant sizes N_A and N_B . In the CM model, introgression is assumed to occur continuously over time from the ancestral split to present. In the AM model, migration is restricted to early times of speciation. In the SC model, migration occurs after an initial period of SI between the two daughter species. For the CM, AM, and SC models, we used the scaled effective migration rates $M = 4Nm$ (M_A is the effective migration rate from *C. intestinalis* B to *C. intestinalis* A, and M_B is the effective migration rate from *C. intestinalis* A to *C. intestinalis* B), where m is the fraction of a population which is made up of effective migrants from the other population at each generation. To statistically evaluate the relative importance of GWH in our observed data set, we compared the two alternative models, that is, “homo” and “hetero,” within the CM, AM, and SC models: in the “homo” version, all loci shared the same effective migration rate m , whereas in the “hetero” version, the m values varied among loci.

For each model, we simulated 8×10^6 multilocus data sets, each composed of 852 loci fitting the length and sampling size of the 852 observed loci. We used large uniform prior distributions for all parameters common to all models after pre-evaluating the relevance of the chosen prior-model combinations according to Cornuet et al. (2010). Prior θ_A/θ_{ref} and θ_B/θ_{ref} distributions were uniform at intervals 0–10 and 0–40, respectively, with $\theta_{ref} = 4N_{ref}\mu$, N_{ref} is the number of effective individuals of the reference population, arbitrarily set at 82,500, and μ is the mutation rate of 1×10^{-8} /bp/generation. This rate was estimated in a phylogenetic study of the tunicate versus vertebrate substitution rate (Tsagkogeorga et al. 2012). We also integrated intralocus recombination at a rate

of 0.7×10^{-8} /bp/generation as estimated (Haubold et al. 2010). The prior distribution for $\theta_{\text{anc}}/\theta_{\text{ref}}$ was uniform over the 0–70 interval. The $T_{\text{split}}/(4N_{\text{ref}})$ ratio was sampled from the interval 0–70 generations, conditioning the parameters T_{AM} and T_{SC} to be uniformly chosen within the 0– T_{split} interval. For the three alternative speciation models with homogeneous migration, a single effective migration rate shared by all loci but different for M_A and M_B was sampled from the uniform distribution 0–15. For the alternative “hetero” models, heterogeneity in effective migration rates among loci was generated by independent samplings among loci from a shared Beta distribution shaped by two parameters, that is, “ a ” (randomly chosen in 0–20) and “ b ” (randomly chosen in 0–200). As the Beta distribution is merely defined on 0–1, we used a multiplier scalar “ c ” randomly chosen in the uniform 0–15 once per multilocus simulation to rescale the genomic distribution of introgression rates between 0 and “ c ”. More precisely, we assume that the Beta distribution shaping m/s (where m is the migration rate and s the selection coefficient of migrant alleles) only reflects genomic variation in s , although we do not know whether selection directly acts at the sequenced loci. Such variation in direct or indirect selection effects can be approximated by coalescent-based simulations modeling variation in effective migration (Barton and Bengtsson 1986). Prior distributions were computed using a modified version of Priorgen software (Ross-Ibarra et al. 2008), and coalescent simulations were run using Msnsam (Ross-Ibarra et al. 2008), a modified version of the Ms program (Hudson 2002), allowing variations in sample size among loci under an infinite site mutation model.

Model Selection

We statistically evaluated alternative speciation models by a hierarchical procedure (Fagundes et al. 2007). First, for each of the CM, AM, and SC models, we evaluated the posterior probabilities of the two alternative Homo and Hetero versions. Second, we separately compared the best version of each of the three models with SI. Finally, we compared the best-supported versions of the three migration models. Posterior probabilities for each candidate model were estimated using a feed-forward neural network implementing a nonlinear multivariate regression by considering the model itself as an additional parameter to be inferred under the ABC framework using the R package “abc” (Csillery et al. 2012). The 0.025% replicate simulations nearest to the observed values for the summary statistics were selected, and these were weighted by an Epanechnikov kernel that peaks when $S_{\text{obs}} = S_{\text{sim}}$. Computations were performed using 50 trained neural networks and 10 hidden networks in the regression.

Models were checked through an analysis of 1,000 pseudo-observed data sets simulated for each compared model, using random parameter values from the same prior distributions as the analysis performed on the observed data set. For each pairwise model comparison, we applied the earlier-described model choice procedure to compute the posterior probabilities of the pseudo-observed data set given a model. The relative probability distributions over the 1,000 replicates were then used to compute the probabilities that the best

supported models are correct given the posterior probabilities obtained for the observed data set (Fagundes et al. 2007; Cornuet et al. 2008).

Parameter Estimation

Posterior distributions of the parameters describing the best model were estimated using a nonlinear regression procedure. Parameters were first transformed according to a log-tangent transformation (Hamilton et al. 2005). Only the 2,000 replicate simulations providing the smallest associated Euclidean distance $\delta = ||S_{\text{obs}} - S_{\text{sim}}||$ were considered. Then the joint posterior parameter distribution was obtained by means of weighted nonlinear multivariate regressions of the parameters on the summary statistics. For each regression, 50 feed-forward neural networks and 20 hidden networks were trained using the R package “abc” (Csillery et al. 2012).

Locus-Specific Model Testing

To gain insight into introgression at the locus scale, we investigated the best-fitting introgression model for each of the 852 surveyed contigs. Four alternative models were first simulated 1.5×10^6 times per contig using parameter values sampled in the joint-posterior distribution for the five parameters common to all loci (N_A , N_B , N_{anc} , T_{split} and T_{SC}) obtained after ABC analysis of the observed data set. Then, we compared these four models in three successive steps: 1) comparison between M0 ($M_A = M_B = 0$) and M3 (M_A and M_B in 0–15); 2) comparison between the best-supported model in the previous step and M2 (M_A in 0–15, $M_B = 0$); and 3) comparison between the best-supported model in the previous step and M1 (M_B in 0–15, $M_A = 0$). The power of this approach was assessed by a simulation analysis of 1,000 pseudo-observed data sets generated under the four different migration models considered in this study.

Sliding-Window Analysis

We used a sliding-window approach to analyze local impoverishment in sequenced loci inferred as M0. The sliding window was wide of five successive loci, and shifted by one locus along chromosomes. To exclude unrealistically large windows due to heterogeneity in genomic distribution of sequenced loci, we rejected windows containing two successive loci more distant than an arbitrary threshold of 150 kb. We defined a window as significantly enriched in introgressed loci if at least three nonM0 loci were found inside as shown after 10,000 permutations (supplementary fig. S5, Supplementary Material online).

Supplementary Material

Supplementary figures S1–S5 and tables S1–S6 are available at *Molecular Biology and Evolution* online (<http://www.mbe.oxfordjournals.org/>, last accessed April 23, 2013).

Acknowledgments

Numerical results presented in this article were carried out using the ISEM computing cluster, and the authors highly appreciate the helpful support of its staff. The authors thank the two anonymous referees for their helpful feedbacks.

They also thank Jonathan Romiguier for his expertise in primate phylogeny. This work was supported by the European Research Council grant ERC PopPhyl 232 971 to N.G. and an Agence National de la Recherche grant ANR-12-BSV7-0011 to N.B.

References

- Abbott R, Albach D, Ansell S, et al. (39 co-authors). 2013. Hybridization and speciation. *J Evol Biol.* 26:229–246.
- Aparicio S, Chapman J, Stupka E, et al. (41 co-authors). 2002. Whole-genome shotgun assembly and analysis of the genome of *Fugu rubripes*. *Science* 297(5585):1301–1310.
- Arnold ML. 2004. Transfer and origin of adaptations through natural hybridization: were Anderson and Stebbins right? *Plant Cell* 16(3):562–570.
- Barton NH. 2001. The evolutionary consequences of gene flow and local adaptation: future approaches. In: Clobert J, Danchin E, Dhondt AA, Nichols JD, editors. *Dispersal*. New York: Oxford University Press. p. 329–340.
- Barton N. 2013. Does hybridisation influence speciation? *J Evol Biol.* 26: 267–269.
- Barton N, Bengtsson BO. 1986. The barrier to genetic exchange between hybridising populations. *Heredity* 57(3):357–376.
- Beaumont MA, Nielsen R, Robert C, et al. (23 co-authors). 2010. In defence of model-based inference in phylogeography. *Mol Ecol.* 19(3):436–446.
- Beaumont MA, Zhang W, Balding DJ. 2002. Approximate Bayesian computation in population genetics. *Genetics* 162(4):2025–2035.
- Becquet C, Przeworski M. 2007. A new approach to estimate parameters of speciation models with application to apes. *Genome Res.* 17(10):1505–1519.
- Bickford D, Lohman DJ, Sodhi NS, Ng PK, Meier R, Winker K, Ingram KK, Das I. 2007. Cryptic species as a window on diversity and conservation. *Trends Ecol Evol.* 22(3):148–155.
- Bierne N. 2010. The distinctive footprints of local hitchhiking in a varied environment and global hitchhiking in a subdivided population. *Evolution* 64(11):3254–3272.
- Bierne N, Bonhomme F, Boudry P, Szulkin M, David P. 2006. Fitness landscapes support the dominance theory of postzygotic isolation in the mussels *Mytilus edulis* and *M. galloprovincialis*. *Proc Biol Sci.* 273:1253–1260.
- Bierne N, Borsa P, Daguin C, Jollivet D, Viard F, Bonhomme F, David P. 2003. Introgression patterns in the mosaic hybrid zone between *Mytilus edulis* and *M. galloprovincialis*. *Mol Ecol.* 12(2):447–461.
- Bikard D, Patel D, Mett  C. Le, Giorgi V, Camilleri C, Bennett MJ, Loudet O. 2009. Divergent evolution of duplicate genes leads to genetic incompatibilities within *A. thaliana*. *Science* 323(5914): 623–626.
- Boon E, Faure MF, Bierne N. 2009. The flow of antimicrobial peptide genes through a genetic barrier between *Mytilus edulis* and *M. galloprovincialis*. *J Mol Evol.* 68(5):461–474.
- Bourelat SJ, Juliusdottir T, Lowe CJ, et al. (14 co-authors). 2006. Deuterostome phylogeny reveals monophyletic chordates and the new phylum Xenoturbellida. *Nature* 444(7115):85–88.
- Brumfield RT, Jernigan RW, McDonald DB, Braun MJ. 2001. Evolutionary implications of divergent clines in an avian (*Manacus*: Aves) hybrid zone. *Evolution* 55(10):2070–2087.
- Butlin RK. 2010. Population genomics and speciation. *Genetica* 138(4):409–418.
- Caputi L, Andreakis N, Mastrototaro F, Cirino P, Vassillo M, Sordino P. 2007. Cryptic speciation in a model invertebrate chordate. *Proc Natl Acad Sci U S A.* 104(22):9364–9369.
- Carling MD, Brumfield RT. 2008. Haldane's rule in an avian system: using cline theory and divergence population genetics to test for differential introgression of mitochondrial, autosomal, and sex-linked loci across the *Passerina* bunting hybrid zone. *Evolution* 62(10): 2600–2615.
- Castric V, Bechsgaard J, Schierup MH, Vekemans X. 2008. Repeated adaptive introgression at a gene under multiallelic balancing selection. *PLoS Genet.* 4(8):e1000168.
- Cianchi R, Ungaro A, Marini M, Bullini L. 2003. Differential patterns of hybridization and introgression between the swallowtails *Papilio machaon* and *P. hospiton* from Sardinia and Corsica islands (Lepidoptera, Papilionidae). *Mol Ecol.* 12(6):1461–1471.
- Clark AG. 1997. Neutral behavior of shared polymorphism. *Proc Natl Acad Sci U S A.* 94(15):7730–7734.
- Cornuet J-M, Ravigné V, Estoup A. 2010. Inference on population history and model checking using DNA sequence and microsatellite data with the software DIYABC (v1.0). *BMC Bioinformatics* 11:401.
- Cornuet J-M, Santos F, Beaumont MA, Robert CP, Marin JM, Balding DJ, Guillemaud T, Estoup A. 2008. Inferring population history with DIY ABC: a user-friendly approach to approximate Bayesian computation. *Bioinformatics* 24(23):2713–2719.
- Csill ry K, Blum MGB, Gaggiotti OE, Fran ois O. 2010. Approximate Bayesian computation (ABC) in practice. *Trends Ecol Evol.* 25(7):410–418.
- Csill ry K, Fran ois O, Blum MGB. 2012. abc: an R package for approximate Bayesian computation (ABC). *Methods Ecol Evol.* 3:475–479.
- Daigle RM, Herbinger CM. 2009. Ecological interactions between the vase tunicate (*Ciona intestinalis*) and the farmed blue mussel (*Mytilus edulis*) in Nova Scotia, Canada. *Aquatic Invasions* 4(1):177–187.
- Dehal P, Satou Y, Campbell RK, et al. (97 co-authors). 2002. The draft genome of *Ciona intestinalis*: insights into chordate and vertebrate origins. *Science* 298(5601):2157–2167.
- Delsuc F, Brinkmann H, Chourrout D, Philippe H. 2006. Tunicates and not cephalochordates are the closest living relatives of vertebrates. *Nature* 439(7079):965–968.
- Dunn MP, Di Gregorio A. 2009. The evolutionarily conserved leprecan gene: its regulation by Brachyury and its role in the developing *Ciona notochord*. *Dev Biol.* 328(2):561–574.
- Fagundes NJR, Ray N, Beaumont M, Neuenschwander S, Salzano FM, Bonatto SL, Excoffier L. 2007. Statistical evaluation of alternative models of human evolution. *Proc Natl Acad Sci U S A.* 104(45):17614–17619.
- Feder JL, Egan SP, Nosil P. 2012. The genomics of speciation-with-gene-flow. *Trends Genet.* 28(7):342–350.
- Fu YX, Li WH. 1993. Statistical tests of neutrality of mutations. *Genetics* 133(3):693–709.
- Gattepaille LM, Jakobsson M, Blum MG. Inferring population size changes with sequence and SNP data: lessons from human bottlenecks. *Heredity*. Advance Access published February 20, 2013, doi: 10.1038/hdy.2012.120.
- Gayral P, Weinert L, Chiari Y, Tsagkogeorga G, Ballenghien M, Galtier N. 2011. Next-generation sequencing of transcriptomes: a guide to RNA isolation in nonmodel animals. *Mol Ecol Resour.* 11(4):650–661.
- Gladieux P, Vercken E, Fontaine MC, Hood ME, Jonot O, Couloux A, Giraud T. 2011. Maintenance of fungal pathogen species that are specialized to different hosts: allopatric divergence and introgression through secondary contact. *Mol Biol Evol.* 28(1):459–471.
- Gladieux P, Zhang X-G, Afoufa-Bastien D, Valdebenito Sanhueza RM, Sbaghi M, Le Cam B. 2008. On the origin and spread of the Scab disease of apple: out of central Asia. *PLoS One* 3(1):e1455.
- Gosset CC, Bierne N. 2013. Differential introgression from a sister species explains high FST outlier loci within a mussel species. *J Evol Biol.* 26: 14–26.
- Goubet PM, Berg s H, Bellec A, et al. (11 co-authors). 2012. Contrasted patterns of molecular evolution in dominant and recessive self-incompatibility haplotypes in *Arabidopsis*. *PLoS Genet.* 8(3): e1002495.
- Hamilton G, Currat M, Ray N, Heckel G, Beaumont M, Excoffier L. 2005. Bayesian estimation of recent migration rates after a spatial expansion. *Genetics* 170(1):409–417.
- Harada Y, Takagaki Y, Sunagawa M, Saito T, Yamada L, Taniguchi H, Shoguchi E, Sawada H. 2008. Mechanism of self-sterility in a hermaphroditic chordate. *Science* 320(5875):548–550.

- Harrison RG. 1993. Hybrids and hybrid zones: historical perspective. In: Harrison RG, editor. Hybrid zones and the evolutionary process. New York: Oxford University Press. p. 3–12.
- Haubold B, Pfaffelhuber P, Lynch M. 2010. mlRho—a program for estimating the population mutation and recombination rates from shotgun-sequenced diploid genomes. *Mol Ecol*. 19(1 Suppl): 277–284.
- Helbig AJ, Salomon M, Bensch S, Seibold I. 2001. Male-biased gene flow across an avian hybrid zone: evidence from mitochondrial and microsatellite DNA. *J Evol Biol*. 14(2):277–287.
- Heliconius Genome Consortium. 2012. Butterfly genome reveals promiscuous exchange of mimicry adaptations among species. *Nature* 487(7405):94–98.
- Hewitt G. 2000. The genetic legacy of the quaternary ice ages. *Nature* 405(6789):907–913.
- Hey J, Nielsen R. 2004. Multilocus methods for estimating population sizes, migration rates and divergence time, with applications to the divergence of *Drosophila pseudoobscura* and *D. persimilis*. *Genetics* 167(2):747–760.
- Hey J, Nielsen R. 2007. Integration within the Felsenstein equation for improved Markov chain Monte Carlo methods in population genetics. *Proc Natl Acad Sci U S A*. 104(8):2785–2790.
- Hey J, Pinho C. 2012. Population genetics and objectivity in species diagnosis. *Evolution* 66(5):1413–1429.
- Hobolth A, Christensen OF, Mailund T, Schierup MH. 2007. Genomic relationships and speciation times of human, chimpanzee, and gorilla inferred from a coalescent hidden Markov model. *PLoS Genet*. 3(2):e7.
- Holland LZ, Gibson-Brown JJ. 2003. The *Ciona intestinalis* genome: when the constraints are off. *Bioessays* 25(6):529–532.
- Hu A, Meehl GA, Otto-Bliesner BL, Waelbroeck C, Han W, Loutre M-F, Lambeck K, Mitrovica JX, Rosenbloom N. 2010. Influence of Bering Strait flow and North Atlantic circulation on glacial sea-level changes. *Nat Geosci*. 3:118–121.
- Hudson RR. 2002. Generating samples under a Wright-Fisher neutral model of genetic variation. *Bioinformatics* 18:337–338.
- Hurt C, Anker A, Knowlton N. 2009. A multilocus test of simultaneous divergence across the Isthmus of Panama using snapping shrimp in the genus *Alpheus*. *Evolution* 63(2):514–530.
- Kim M, Cui M-L, Cubas P, Gillies A, Lee K, Chapman MA, Abbott RJ, Coen E. 2008. Regulatory genes control a key morphological and ecological trait transferred between species. *Science* 322(5904):1116–1119.
- Kliman RM, Andolfatto P, Coyne JA, Depaulis F, Kreitman M, Berry AJ, McCarter J, Wakeley J, Hey J. 2000. The population genetics of the origin and divergence of the *Drosophila simulans* complex species. *Genetics* 156(4):1913–1931.
- Knowlton N, Weigt LA. 1998. New dates and new rates for divergence across the Isthmus of Panama. *Proc R Soc Lond B Biol Sci*. 265(1412):2257–2263.
- Knowlton N, Weigt LA, Solórzano LA, Mills DK, Bermingham E. 1993. Divergence in proteins, mitochondrial DNA, and reproductive compatibility across the isthmus of Panama. *Science* 260(5114):1629–1632.
- Lambert CC, Lambert G. 1998. Non-indigenous ascidians in southern California harbors and marinas. *Marine Biol*. 130(4):675–688.
- Lawniczak MKN, Emrich SJ, Holloway AK, et al. (30 co-authors). 2010. Widespread divergence between incipient *Anopheles gambiae* species revealed by whole genome sequences. *Science* 330(6003):512–514.
- Lessios HA. 2008. The Great American Schism: divergence of marine organisms after the rise of the central American Isthmus. *Annu Rev Ecol Syst*. 39:63–91.
- Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25(14): 1754–1760.
- Liu F, Van der Lijn F, Schurmann C, et al. (34 co-authors). 2012. A genome-wide association study identifies five loci influencing facial morphology in Europeans. *PLoS Genet*. 8(9):e1002932.
- Llopart A, Lachaise D, Coyne JA. 2005. Multilocus analysis of introgression between two sympatric sister species of *Drosophila*: *Drosophila yakuba* and *D. santomea*. *Genetics* 171(1):197–210.
- Mallet J. 2007. Hybrid speciation. *Nature* 446(7133):279–283.
- Masley JP, Presgraves DC. 2007. High-resolution genome-wide dissection of the two rules of speciation in *Drosophila*. *PLoS Biol*. 5(9):e243.
- Matute DR, Butler IA, Turissini DA, Coyne JA. 2010. A test of the snowball theory for the rate of evolution of hybrid incompatibilities. *Science* 329(5998):1518–1521.
- Michel AP, Sim S, Powell TH, Taylor MS, Nosil P, Feder JL. 2010. Widespread genomic divergence during sympatric speciation. *Proc Natl Acad Sci U S A*. 107:9724–9729.
- Moyle LC, Nakazato T. 2010. Hybrid incompatibility “snowballs” between *Solanum* species. *Science* 329(5998):1521–1523.
- Mullen SP, Dopman EB, Harrison RG. 2008. Hybrid zone origins, species boundaries, and the evolution of wing-pattern diversity in a polytypic species complex of North American admiral butterflies (Nymphalidae: Limenitis). *Evolution* 62(6):1400–1417.
- Nielsen R, Wakeley J. 2001. Distinguishing migration from isolation: a Markov chain Monte Carlo approach. *Genetics* 158(2):885–896.
- Nydam M, Harrison R. 2007. Genealogical relationships within and among shallow-water *Ciona* species (Ascidacea). *Marine Biol*. 151(5):1839–1847.
- Nydam ML, Harrison RG. 2010. Polymorphism and divergence within the ascidian genus *Ciona*. *Mol Phylogenet Evol*. 56(2):718–726.
- Nydam ML, Harrison RG. 2011. Introgression despite substantial divergence in a broadcast spawning marine invertebrate. *Evolution* 65(2):429–442.
- Orr HA. 1995. The population genetics of speciation: the evolution of hybrid incompatibilities. *Genetics* 139(4):1805–1813.
- Pardo-Diaz C, Salazar C, Baxter SW, Merot C, Figueiredo-Ready W, Joron M, McMillan WO, Jiggins CD. 2012. Adaptive introgression across species boundaries in *Heliconius* butterflies. *PLoS Genet*. 8(6):e1002752.
- Patterson N. 2006. Genetic evidence for complex speciation of humans and chimpanzees. *Nature* 441:1103–1108.
- Perelman P, Johnson WE, Roos C, et al. (14 co-authors). 2011. A molecular phylogeny of living primates. *PLoS Genet*. 7(3):e1001342.
- Presgraves DC. 2003. A fine-scale genetic analysis of hybrid incompatibilities in *Drosophila*. *Genetics* 163(3):955–972.
- Pritchard JK, Seielstad MT, Perez-Lezaun A, Feldman MW. 1999. Population growth of human Y chromosomes: a study of Y chromosome microsatellites. *Mol Biol Evol*. 16(12):1791–1798.
- Ramírez-Soriano A, Ramos-Onsins SE, Rozas J, Calafell F, Navarro A. 2008. Statistical power analysis of neutrality tests under demographic expansions, contractions and bottlenecks with recombination. *Genetics* 179(1):555–567.
- Ramos-Onsins SE, Rozas J. 2002. Statistical properties of new neutrality tests against population growth. *Mol Biol Evol*. 19(12):2092–2100.
- Ramos-Onsins SE, Stranger BE, Mitchell-Olds T, Aguadé M. 2004. Multilocus analysis of variation and speciation in the closely related species *Arabidopsis halleri* and *A. lyrata*. *Genetics* 166(1): 373–388.
- Ramsay A, Davidson J, Landry T, Arsenault G. 2008. Process of invasiveness among exotic tunicates in Prince Edward Island, Canada. *Biol Invasions*. 10(8):1311–1316.
- Robinson TB, Griffiths CL, McQuaid C, Rius M. 2005. Marine alien species of South Africa—status and impacts. *African J Marine Sci*. 27(1):297–306.
- Ross-Ibarra J, Tenaillon M, Gaut BS. 2009. Historical divergence and gene flow in the genus *Zea*. *Genetics* 181(4):1399–1413.
- Ross-Ibarra J, Wright SI, Foxe JP, Kawabe A, DeRose-Wilson L, Gos G, Charlesworth D, Gaut BS. 2008. Patterns of polymorphism and demographic history in natural populations of *Arabidopsis lyrata*. *PLoS One* 3(6):e2411.
- Roux C, Castric V, Pauwels M, Wright SI, Saumitou-Laprade P, Vekemans X. 2011. Does speciation between *Arabidopsis halleri* and *Arabidopsis lyrata* coincide with major changes in a molecular target of adaptation? *PLoS One* 6(11):e26872.

- Roux C, Pauwels M, Ruggiero M-V, Charlesworth D, Castric V, Vekemans X. 2012. Recent and ancient signature of balancing selection around the S-locus in *Arabidopsis halleri* and *A. lyrata*. *Mol Biol Evol.* 30: 435–447.
- 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 Biol.* 159(7):1611–1619.
- Schierup MH, Charlesworth D, Vekemans X. 2000. The effect of hitchhiking on genes linked to a balanced polymorphism in a subdivided population. *Genet Res.* 76(1):63–73.
- Sjödén P, Sjöstrand AE, Jakobsson M, Blum MGB. 2012. Resequencing data provide no evidence for a human bottleneck in Africa during the penultimate glacial period. *Mol Biol Evol.* 29(7):1851–1860.
- Soltis PS, Soltis DE. 2009. The role of hybridization in plant speciation. *Annu Rev Plant Biol.* 60:561–588.
- Song Y, Endepols S, Klemann N, Richter D, Matuschka FR, Shih CH, Nachman MW, Kohn MH. 2011. Adaptive introgression of anticoagulant rodent poison resistance by hybridization between old world mice. *Curr Biol.* 21(15):1296–1301.
- St Onge KR, Foxe JP, Li J, Li H, Holm K, Corcoran P, Slotte T, Lascoux M, Wright SI. 2012. Coalescent-based analysis distinguishes between allo- and autopolyploid origin in Shepherd’s Purse (*Capsella bursa-pastoris*). *Mol Biol Evol.* 29(7):1721–1733.
- Städler T, Haubold B, Merino C, Stephan W, Pfaffelhuber P. 2009. The impact of sampling schemes on the site frequency spectrum in nonequilibrium subdivided populations. *Genetics* 182(1):205–216.
- Sun Y, Corcoran P, Menkis A, Whittle CA, Andersson SGE, Johannesson H. 2012. Large-scale introgression shapes the evolution of the mating-type chromosomes of the filamentous ascomycete *Neurospora tetrasperma*. *PLoS Genet.* 8(7):e1002820.
- Suzuki MM, Nishikawa T, Bird A. 2005. Genomic approaches reveal unexpected genetic divergence within *Ciona intestinalis*. *J Mol Evol.* 61(5):627–635.
- Tajima F. 1983. Evolutionary relationship of DNA sequences in finite populations. *Genetics* 105(2):437–460.
- Tajima F. 1989. The effect of change in population size on DNA polymorphism. *Genetics* 123(3):597–601.
- Tao Y, Chen S, Hartl DL, Laurie CC. 2003. Genetic dissection of hybrid incompatibilities between *Drosophila simulans* and *D. mauritiana*. I. Differential accumulation of hybrid male sterility effects on the X and autosomes. *Genetics* 164(4):1383–1397.
- Tavaré S, Balding DJ, Griffiths RC, Donnelly P. 1997. Inferring coalescence times from DNA sequence data. *Genetics* 145(2):505–518.
- Templeton A. 2006. Population genetics and microevolutionary theory. Hoboken (NJ): John Wiley & Sons.
- 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 Biol Evol.* 4(8):740–749.
- Turelli M, Orr HA. 1995. The dominance theory of Haldane’s rule. *Genetics* 140:389–402.
- Turelli M, Orr HA. 2000. Dominance, epistasis and the genetics of post-zygotic isolation. *Genetics* 154:1663–1679.
- Turner TL, Hahn MW, Nuzhdin SV. 2005. Genomic islands of speciation in *Anopheles gambiae*. *PLoS Biol.* 3(9):e285.
- Via S. 2009. Natural selection in action during speciation. *Proc Natl Acad Sci U S A.* 106(Suppl): 9939–9946.
- Wakeley J, Hey J. 1997. Estimating ancestral population parameters. *Genetics* 145(3):847–855.
- Watterson G. 1975. On the number of segregating sites in genetical models without recombination. *Theor Popul Biol.* 7(2):256–276.
- Whitney KD, Baack EJ, Hamrick JL, et al. (11 co-authors). 2010. A role for nonadaptive processes in plant genome size evolution? *Evolution* 64(7):2097–2109.
- Whitney KD, Randell RA, Rieseberg LH. 2006. Adaptive introgression of herbivore resistance traits in the weedy sunflower *Helianthus annuus*. *Am Nat.* 167(6):794–807.
- Won Y-J, Hey J. 2005. Divergence population genetics of chimpanzees. *Mol Biol Evol.* 22(2):297–307.
- Wu C-I. 2001. The genic view of the process of speciation. *J Evol Biol.* 14(6):851–865.
- Zhan A, Macisaac HJ, Cristescu ME. 2010. Invasion genetics of the *Ciona intestinalis* species complex: from regional endemism to global homogeneity. *Mol Ecol.* 19(21):4678–4694.