Received Date : 08-May-2014
Revised Date : 03-Sep-2014
Accepted Date : 10-Sep-2014
Article type : Original Article

Wheat alleles introgress into selfing wild relatives: empirical estimates from Approximate Bayesian Computation in *Aegilops triuncialis*.

Authors

Mila Pajkovic[1]*, Sylvain Lappe[2]*, Rachel Barman[1]*, Christian Parisod[2], Samuel Neuenschwander[1,3], Jerome Goudet[1,3], Nadir Alvarez[1], Roberto Guadagnuolo[2], Francois Felber[4] and Nils Arrigo[1,2†]

[1]Department of Ecology and Evolution, Institute of Biology, University of Lausanne, Biophore Dorigny, CH-1015 Lausanne, Switzerland

[2]Laboratory of Evolutionary Botany, Institute of Biology, University of Neuchatel, Emile-Argand 11, CH-2001 Neuchatel, Switzerland

[3]Vital-IT, Swiss Institute of Bioinformatics, University of Lausanne, Switzerland

[4]Musée et Jardins botaniques cantonaux vaudois, CH-1007 Lausanne, Switzerland

*These authors have contributed equally to this work.

[†]Corresponding author

Keywords

Crop to wild gene flow, autogamy, admixture, ABC, transgene escape, containment.

Corresponding author

Nils Arrigo

Department of Ecology and Evolution, Institute of Biology, University of Lausanne, Biophore Dorigny, CH-1015 Lausanne, Switzerland

Fax number: +41 21 692 4165

E-mail: nils.arrigo@unil.ch

Running title

An ABC perspective on crop to wild gene flow

Abstract

Extensive gene flow between wheat (*Triticum* sp.) and several wild relatives of the genus *Aegilops* has recently been detected despite notoriously high levels of selfing in these species. Here, we assess and model the spread of wheat alleles into natural populations of the barbed goatgrass (*Aegilops triuncialis)*, a wild wheat relative prevailing in the Mediterranean flora. Our sampling, based on an extensive survey of 31 *Ae. triuncialis* populations collected along a 60 km x 20 km area in Southern Spain (Grazalema Mountain chain, Andalousia, totaling 458 specimens) is completed with 33 wheat cultivars representative of the European domesticated pool. All specimens were genotyped with amplified fragment length polymorphism with the aim of estimating wheat admixture levels in *Ae. triuncialis* populations. This survey first confirmed extensive hybridization and backcrossing of wheat into the wild species. We then used explicit modeling of populations and Approximate Bayesian Computation to estimate the selfing rate of *Ae. triuncialis* along with the magnitude, the tempo and the geographic distance over which wheat alleles introgress into *Ae. triuncialis* populations. These simulations confirmed that extensive introgression of wheat alleles (2.7×10^{-4} wheat immigrants for each *Ae. triuncialis* resident, at each generation) into *Ae. triuncialis* occurs despite a high selfing rate (Fis ≈ 1 and selfing rate = 97%). These results are discussed in light of risks associated with the release of genetically-modified wheat cultivars in Mediterranean agrosystems.

Introduction

Hybridization, the interbreeding of different species, has important evolutionary consequences by promoting the transfer of genetic diversity among species via introgression (Abbott *et al.* 2013). Such exchanges have prompted important

research efforts, especially since the release of genetically engineered organisms has required a thorough survey of wild species into which crop (trans)genes might escape. Accordingly, the introgression of crop genes is well documented for wild outcrossing species and has already resulted in severe production and biosafety concerns (e.g. in *Amaranthus*, *Beta*, *Helianthus*, *Oryza*, *Brassica*, *Raphanus*, *Gossypium*, *Zea*, *Lactuca*, *Prunus*, *Agrostis*, Tranel *et al.* 2002; Arnaud *et al.* 2003; Massinga *et al.* 2003; Lu & Snow 2005; FitzJohn *et al.* 2007; Warwick *et al.* 2008; Trucco *et al.* 2009; Snow *et al.* 2010; Wegier *et al.* 2011; van Heerwaarden *et al.* 2012; Uwimana *et al.* 2012; Delplancke *et al.* 2012; Snow 2012).

Here, we focus on wheat (*Triticum* sp. L.), a worldwide leading crop (671 million tons produced worldwide in 2012, www.faostat.fao.org), to assess to what extent selfing species are at risk and deserve as much attention as that given to outcrossers (but see Lu & Snow, 2005). Wheat and most of its wild counterparts (and ancestors) from *Aegilops* L.*,* a genus of ca. 40 species, are renowned selfers with floral adaptations enforcing self-pollination (i.e. cleistogamy, where the pollen is released while the spikelets are still closed, reviewed in Kilian *et al.* 2011, see also Zaharieva & Monneveux 2006). Wheat is frequently grown in areas where *Aegilops* species co-occur. In addition, both genera belong to a diploid / polyploid complex of closely related species (Fan *et al.* 2013) that are interfertile to some extent. This genetic relatedness and geographical proximity has been widely exploited for wheat improvement purposes and sustains occasional wheat x *Aegilops* hybrids in natural populations (Hegde & Waines 2004; Loureiro *et al.* 2006, 2007a; 2007b, 2009; Zaharieva & Monneveux 2006). For instance, bread wheat (*T. aestivum*) resulted from the spontaneous hybridization of durum wheat (*T. durum*) with *Ae. tauschii* (Feldman & Levy 2009). There is at least one study documenting long standing

signatures of wheat introgression in populations of *Aegilops* (Weissmann *et al.* 2005), but most authors assume that reproductive barriers lead to limited genetic exchanges between these genera (Waines & Hegde 2003; Zaharieva & Monneveux 2006; Felber *et al.* 2007). As a result, wheat is generally considered a "moderate risk crop" for (trans-)gene escape (Eastham & Sweet 2002; Stewart *et al.* 2003; Felber *et al.* 2007).

Several recent molecular surveys have however questioned the strength of reproductive barriers, by revealing the pervasive introgression of wheat DNA in natural populations of the Mediterranean *Ae. triuncialis* L. (Arrigo *et al.* 2011; Parisod *et al.* 2013). However, these studies were mostly pattern-based and had limited insight into the demographic parameters associated with the introgression process. Indeed, the efficiency of selfing, the effective immigration rate of wheat alleles into *Aegilops* populations, and the geographical extent over which introgression takes place are still poorly documented. These parameters are of major importance for risk assessment purposes and ask for thorough estimations.

Here, we aim at bridging this gap by analyzing a representative set of natural *Ae. triuncialis* populations, collected in areas where introgression of wheat alleles has been documented (Arrigo *et al.* 2011, although this study considered a larger sampling area), and explicitly addressing the demographic parameters sustaining the ongoing introgression process. We rely on genome-wide amplified fragment length polymorphism fingerprints (AFLP), a cost efficient and reliable method to record the introgression of wheat DNA into *Ae. triuncialis* (Arrigo *et al.* 2011). We then use population modeling coupled to Approximate Bayesian Computation (Beaumont *et al.* 2002; Csilléry *et al.* 2010; Beaumont 2010; Sunnaker *et al.* 2013), to explore how alternative introgression scenarios fit our empirical observations and estimate the

key demographic parameters under which wheat alleles introgress into natural populations of *Ae. triuncialis*.

Materials and methods

Sampling

A total of 458 specimens of *Ae. triuncialis* were collected along a 60 x 20 km area located in the Grazalema National Park (36°45'N - 5°21'W, Andalusia, Southern Spain). The sampling included 31 natural populations (15 specimens / population on average) growing at increasing distances from cultivated fields (ranging from 0 m to 8.5 km, Fig. 1 and Table 1). The specimens were collected during April 2008 and identified morphologically (following van Slageren 1994), using the maternal spikes from which the collected plantlets were emerging. The sampling was completed with 33 representative wheat reference specimens (19 bread wheat and 14 durum wheat) that were either collected in Andalusia or provided by the seed collection of the Agroscope Changins Waedenswil agronomic research station (Table S1).

AFLP fingerprinting

DNA extractions were performed from 10 mg of silica-gel dried leaves, using a CTAB protocol (Chen & Ronald 1999). DNA concentrations were standardized at 10 ng/µl. All specimens were analyzed with AFLP, following Gugerli *et al.* 2008. The PCR reactions were conducted on a PTC-100 thermocycler, with 96-well plates where specimens were randomly distributed. Two selective primer pairs were used (FAM-*Eco*RI-ATC / *Mse*I-CAA and NED-*Eco*RI-ACT / *Mse*I-CTG). PCR products were analyzed with an ABI 3730XL capillary sequencer (Applied Biosystems, Foster City, CA, USA, service provided by Macrogen Inc. Seoul, Korea). The detection and size calculation of AFLP amplicons was performed from raw electropherograms with

PeakScanner V1.0 (ABI), using the default peak detection parameters, except a light peak smoothing. The scoring (recording of the presence / absence of AFLP bands for each sample) was performed with RawGeno V2.0-1, an automated scoring R CRAN package (Arrigo *et al*. 2009) and following recommendations of Arrigo *et al.* 2012: scoring range = 100 - 400 bp, minimum bin width = 1 bp, maximum bin width = 2 bp, minimum fluorescence intensity = 100 rfu, minimum reproducibility threshold = 85%. Seven control specimens and one blank were amplified on each plate, in a way to assess between-plate variability and PCR template purity. Nine samples chosen randomly from each plate (representing 10% of the final dataset) were also replicated on another PCR plate to identify non-reproducible bands. The average error rate of AFLP reactions across all replicated specimens was of 6%.

Diversity and F-statistics

Genetic diversity and differentiation statistics were computed in order to describe the empirical dataset. The percentage of polymorphic AFLP markers per population was estimated with custom R CRAN scripts (v. 3.0.2; scripts available upon request to the last author). We then computed F-statistics with ABC4F (Foll *et al.* 2008), an algorithm co-estimating inbreeding coefficients (FIS) and population-specific fixation indices (FST) using dominant markers. In contrast to other algorithms, ABC4F skips the estimation of allelic frequencies (which generally relies on Hardy-Weinberg equilibrium - HWE - assumptions, e.g. Falush *et al.* 2007) and treats FIS and FST as model parameters to be directly estimated with Approximate Bayesian Computation (hereafter ABC). This is suitable because our focal species as well as most *Aegilops* species are selfers (van Slageren 1994) and HWE assumptions are likely violated. The algorithm also corrects for ascertainment biases introduced by post-scoring

treatment of AFLP (i.e. removal of non-polymorphic bands) and provides accurate estimates of F-statistics from dominant markers.

Estimations of genetic admixture

Due to HWE deviations (see above), we estimated the admixture of wheat genotypes into *Aegilops* specimens using methods implemented in InStruct v. 1.0 (Gao *et al.* 2007). The approach considers a model treating the population origin of specimens, their admixture level and the global inbreeding level as parameters to be estimated. The model explicitly accounts for departures from HWE equilibrium caused by selfing as well as for allopolyploidy. The computations were performed using the default settings, except assuming allopolyploidy. The parameter estimation relied on four MCMC runs of 500,000 iterations each. The parameters were estimated from data points collected each 1,000 iterations. The first 100 data collections of each run were discarded as a burn-in phase.

Second, we estimated admixture with fuzzy c-means clustering (hereafter FCM), a distance-based algorithm working without explicit biological assumptions. Such non-model based approaches have already been successfully applied in phylogeography and gene flow detection in wheat and *Aegilops* species (Arrigo *et al.* 2010, 2011). The FCM algorithm uses an iterative procedure to (i) assign specimens into a predefined number of groups (i.e. *K*), so that the intra-group variance is minimized and (ii) compute the membership of specimens to each of the *K* groups. As a result, each specimen belongs to some extent to each group, a feature reflected by membership values (e.g. see Gompert *et al.* 2010). These can then be understood as group assignment probabilities and are used here as proxies of admixture levels (as in Arrigo *et al.* 2011). Finally, the algorithm relies on a fuzzification parameter (*r*,

ranging between one and infinity) acting on the stringency of clustering. This parameter must be understood as the "sensitivity" of the analysis and requires a dataset-specific optimization: using $r \approx 1$ produces non-fuzzy clusters (i.e. memberships take either 1 or 0 as values, an outcome equivalent to classical K-means) while larger r values increase the fuzziness of clustering until all individuals are equally assigned to all genetic groups. We tested r values ranging between 1 and 2, all trials provided congruent results (see Supplementary Material 1) and we report here admixture measures obtained with $r = 1.25$. The computations were performed using custom R CRAN scripts (available on request to the last author). The obtained results were used to validate the admixture estimates obtained with InStruct.

All admixture estimations were performed by assuming $K = 2$ groups and the assignment probabilities to the wheat pool were used as a proxy for admixture levels. These measures were summarized at the population level and plotted on a geographic map using ArcMap v. 9.3.1 (ESRI 2009).

Explicit population models and Approximate Bayesian Computation

We were interested in (i) assessing how alternative introgression scenarios fitted our empirical observations and (ii) quantifying the selfing rate of *Ae. triuncialis*, the migration rate and the spatial and temporal spread of wheat alleles into wild *Ae. triuncialis* populations. To this end, we used custom scripts (available on Dryad, doi:10.5061/dryad.7np24) that coupled explicit simulation of populations (using QuantiNEMO v. 1.6.0., Neuenschwander *et al.* 2008) with Approximate Bayesian Computation (Beaumont *et al.* 2002; Csilléry *et al.* 2010; Beaumont 2010; Sunnaker *et al.* 2013). The complete set of models, their implementation, the model selection

procedure, the parameter estimations and the cross-validation of results are explained in the following paragraphs and are further detailed in Supplementary Material 1 and 2.

We explored six introgression scenarios (Fig. 2a) that assessed i) how wheat alleles disperse in the landscape and ii) whether secondary dispersal of wheat alleles occurs within the *Ae. triuncialis* metapopulation. Models 1 and 2 reflected the absence of gene flow between wheat and *Ae. triuncialis*. Models 3 and 4 assumed that wheat to *Aegilops* gene flow occurred with the same probability among all *Ae. triuncialis* populations (i.e. absence of a spatial structure in the occurrence of admixture). Finally, models 5 and 6 accounted for the distance of cultivated fields in the dispersal of wheat alleles (i.e. wheat to *Ae. triuncialis* migration decreasing with distance to cultivated fields). In models 3 and 5, wheat alleles were immigrating into *Ae. triuncialis* only via primary dispersal (i.e. only migration events from the wheat to the *Ae. triuncialis* populations were allowed, no further migration events could occur among *Ae. triuncialis* populations). In contrast, the models 4 and 6 accounted for primary and secondary dispersal events (i.e. migrations among *Ae. triuncialis* populations were allowed).

Model parameters and priors

All models considered one wheat and 31 *Ae. triuncialis* populations, respectively (Fig. 2a and b, Supplementary Material 1). The wheat population was of fixed carrying capacity (N_{wheat} = 5,000 individuals) and acted as a constant supplier of outmigrants. Each *Ae. triuncialis* population could receive wheat immigrants (hereafter $M_{wheat\text{-}to\text{-}Aegilops}$) while migration from *Ae. triuncialis* to the wheat population was not allowed, as this was not of relevance due to annual harvesting of wheat

populations. The number of wheat immigrants reaching each of the 31 *Ae. triuncialis* populations was further adjusted according to a dispersal probability (i.e. $M_{wheat\text{-}to\text{-}Aegilops\_i} = M_{wheat\text{-}to\text{-}Aegilops} \times P_{dispersal\_i}$). Dispersal probabilities ($P_{dispersal\_i}$) were either all equal (models 1 to 4) or varied according to the actual geographic distance isolating each corresponding empirical *Ae. triuncialis* population from the wheat fields (models 5 and 6). In the latter models, $P_{dispersal\_i}$ was determined with a dispersal kernel (a negative exponential function defined in Austerlitz *et al.* (2004) that introduced a shape parameter (hereafter *c,* reflecting the dispersal curve steepness, see Fig. 2c). The same dispersal kernel was also used in models accounting for secondary dispersal (i.e. in models 2, 4 and 6, $M_{wheat\text{-}to\text{-}Aegilops}$ and $M_{Aegilops\text{-}to\text{-}Aegilops}$ were set using the same shape parameter). The dispersal probabilities were rescaled to sum to one, so that all wheat outmigrants were allocated in one or another of the *Ae. triuncialis* populations. We then used an ABC framework to estimate the carrying capacity of the *Ae. triuncialis* populations (prior: $1 < N_{Ae.triuncialis} < 5{,}000$), the selfing rate ($0 < S_{Ae.triuncialis} < 1$), the proportion of wheat outmigrants ($0 < M_{wheat\text{-}to\text{-}Aegilops} < 0.5$), the shape of the dispersal kernel ($0 < c < 15{,}000$, for models 5 and 6) and the number of generations ($1 < N_{gen} < 300$) necessary to produce the empirical admixture levels. All parameters followed uniform prior distributions ($N_{Ae.triuncialis}$, $S_{Ae.triuncialis}$, $M_{wheat\text{-}to\text{-}Aegilops}$ and *c*), except N_{gen} that was sampled from a Beta distribution, with the expectation that most detected introgression events would be of relatively recent origin (see details in Supplementary Material 1). $N_{Ae.triuncialis}$ and $S_{Ae.triuncialis}$ were shared among all *Ae. triuncialis* populations. The simulations were based on 100 biallelic loci that were recoded following a dominant scheme (dominant homozygotes and heterozygotes were recoded as “present” phenotypes, while recessive homozygotes were recoded as “absent”). The mutation rate was fixed at

10^{-7} allele changes per generation and per locus (following Lynch 2010). The simulations were initiated with population-level allele frequencies, estimated from empirical AFLPs after removal of admixed specimens. This strategy allowed initiating simulations with populations that were differentiated in a realistic way and, more importantly, that were free of any recent wheat admixture (see Supplementary Material 1 for further details; this strategy also corrects for any signatures left by ancient admixture events in the sampled populations). Initial wheat genotypes were also simulated from empirical allele frequencies (based on the 33 reference wheat specimens).

Summary statistics, model selection and parameter estimations using ABC

Summary statistics were used to compare the simulation outcomes to the empirical dataset, in a way to select the most likely scenario and estimate its parameters. Here, these statistics were computed by resampling specimens from each simulated population, in a way to account for sampling effort variations in the empirical dataset (see Table 1). Two sets of summary statistics were collected. First, we tracked frequency increases of wheat-diagnostic markers into *Aegilops* populations (i.e. the initial allele frequencies were set so that the "present" AFLP phenotypes would prevail in the wheat population, see Supplementary Material 1). Populations experiencing introgression thus showed an increased average frequency of wheat diagnostic markers. Second, we used fuzzy c-means clustering to gather individual estimates of admixture (c-means was set as explained earlier). Both summary statistics were then summarized at the population level using mean and standard-deviation, resulting in 124 statistics collected per simulation (i.e. 31 populations investigated, each with two admixture estimates, and summarized with mean and

standard-deviation). The same summary statistics were also estimated from the empirical AFLP dataset.

The accuracy of simulations was measured as their Euclidean distance to the empirical summary statistics. On this basis, we could explore how each of the six introgression scenarios (i.e. models 1 to 6) fitted with the empirical observations. The model selection procedure used methods implemented in the "abc" R package (Csilléry *et al.* 2012) and relied on 200,000 simulations per model. We adopted a hierarchical approach, based on pairwise model comparisons, in order to evaluate the respective importance of secondary dispersal and that of the various modes of wheat introgression (see details in Supplementary Material 2). Each comparison relied on posterior probabilities - P(model | empirical data) - computed using the 1,000 best simulations produced by the compared models. The selection procedure was then validated with cross-validations, in a way to estimate the probability of selecting the wrong model associated to a given posterior probability (see details in Supplementary Material 2).

Model parameters (i.e. $N_{Ae.triuncialis}$, $S_{Ae.triuncialis}$, $M_{\text{wheat-to-}Aegilops}$, c and N_{gen}) were estimated by adding another 1,000,000 simulations to the best model. Parameter values were estimated from posterior distributions, computed with the 3,000 best simulations (i.e. 0.25%), out of a total of 1,200,000. These were further refined using neural-net local adjustment implemented in the "abc" R CRAN package (using logit transformation and 50 iterations for the neural-net). Here as well, cross-validations were used to assess the consistency and accuracy of estimations (see details in Supplementary Material 2). The simulations were performed on the Vital-IT High Performance Computing Center.

Results

Diversity and inbreeding coefficients

A total of 491 AFLP bands were produced (on average 145 bands per specimen), of which 188 (38%) and 247 (50%) were polymorphic in the *Ae. triuncialis* and the wheat specimen pools, respectively. The surveyed populations yielded on average 34% ± 4% SE of polymorphic loci (three populations, 824598, 755237 and 789353 were however especially diverse with estimates higher than 40%). The F-statistics estimated by ABC4F revealed high inbreeding levels (average FIS = 0.99 ± 0.02 SE) and a strong population differentiation (average FST = 0.71 ± 0.14 SE) among *Ae. triuncialis* populations (Table 1).

Estimates of admixture levels

Both InStruct and fuzzy c-means assigned wheat and *Ae. triuncialis* specimens to distinct genetic groups and allowed a consistent estimation of genetic admixture levels (Supplementary Material 1). Most *Ae. triuncialis* specimens could be assigned unambiguously to the *Aegilops* group (i.e. 90% of the *Ae. triuncialis* specimens had assignment probabilities to the *Aegilops* group greater than 0.9) and were considered as non-admixed genotypes. The remainder of the surveyed specimens showed increased relatedness to wheat genotypes and were treated here as putatively admixed. These specimens occurred exclusively in populations located close to cultivated fields (e.g. population 824598) and there was a negative (but non-significant) correlative trend between admixture levels and the geographic distance of *Ae. triuncialis* specimens to cultivated fields (Fig. 1).

Model selection and parameter estimation using ABC

The six alternative introgression scenarios (Fig. 2a, Table 2) revealed that including secondary dispersal events within the *Ae. triuncialis* metapopulation decreased the accuracy of simulations (posterior probability of 0.02). This result was supported by cross-validations (p-value = 0.02) and we therefore excluded models 2, 4 and 6. In contrast, accounting for interspecific gene flow dramatically improved our models (i.e. rejection of model 1, with posterior probabilities of 1 and cross-validated p-values of 0, see Supplementary Material 2). Discriminating the remaining scenarios (models 3 and 5) proved more difficult: although model 5 fitted best with empirical observations (see additional details in Supplementary Material 2), model 3 yielded non-negligible posterior probability (0.32). Accordingly, cross-validations confirmed that empirical observations could not allow distinguishing between the two models with significant support (p-value = 0.33). Adding another 1,000,000 simulations to each model did not improve results (data not shown) and it remains unclear whether wheat alleles spread into the *Ae. triuncialis* metapopulation according to a spatially explicit process.

Demographic parameters were estimated using models 3 and 5. Both models provided similar and essentially unbiased results (Fig. 3 and Supplementary Material 2), except that model 5 resulted in wider confidence intervals and the dispersal kernel was not included in model 3. First, both models revealed that $N_{gen} \approx 18$ generations of ongoing wheat immigration (Fig. 3a, widest confidence interval = 2 – 126) were needed to recover the empirical signatures of introgression. The selfing rate of *Ae. triucialis* was high (Fig. 3b), with a posterior distribution mode at $S_{Ae.triuncialis} \approx 0.97$ (widest CI = 0.67 – 1.00). The wheat immigration rate (Fig. 3c), estimated as the number of wheat immigrants per generation, relative to the total

carrying capacity of the *Ae. triuncialis* metapopulation (i.e. jointly estimated from the $N_{Ae.triuncialis}$ and the $M_{wheat\text{-}to\text{-}Aegilops}$ parameters, see details in Fig 2) was high as well, with $N_{immig} \approx 2.7 \times 10^{-4}$ (Fig. 3c, widest CI = 7.4×10^{-5} – 1.1×10^{-3}). The shape of the dispersal kernel proved difficult to estimate, with $c \approx 5230$ (Fig. 3d, widest CI = 756 – 13789). Finally, these estimates were all robust to cross-validations (Supplementary Material 2).

Discussion

Crop to wild gene flow in the face of selfing

The present study detected unambiguous evidence of genetic admixture between wheat and *Ae. triuncialis* (Fig. 1). These results were supported by three independent computational approaches (InStruct and c-means estimations of admixture, along with ABC model selections) and corroborated a growing body of empirical evidence. Accordingly, hybridization and introgression between wheat and *Ae. triuncialis* (among other *Aegilops* species, David *et al.* 2004; Zaharieva & Monneveux 2006), have been reported from botanical surveys (van Slageren 1994), open pollination experiments in field conditions (Loureiro *et al.* 2007a) and population genetic surveys (Arrigo *et al.* 2011; Parisod *et al.* 2013). In addition, our study outlined a predominantly selfing mating system in *Ae. triuncialis*, by revealing strong inbreeding coefficients and high selfing rates. Here as well, our results were in line with the limited pollen production and cleistogamous flowers typical of most selfing *Aegilops* species (Hammer 1980). It is therefore clear that introgression also occurs between domesticated and wild species that are isolated by prezygotic reproductive barriers. Similar transfers have already been reported in some other typically selfing crop-wild species complexes (e.g. Cotton, Wegier *et al.* 2011; Rice,

Lu and Yang 2009; Lettuce, D'Andrea *et al*. 2008). Yet, this route of crop (trans-) gene escape into natural populations has remained largely overlooked by current risk assessment studies (but see Stewart *et al*. 2003; Darmency *et al*. 2007; Felber *et al*. 2007).

Quantitative insights of wheat DNA introgression in Ae.triuncialis

Our AFLP survey suggests that at least 10% of the analyzed *Ae. triuncialis* specimens (representing 45 specimens across 31 populations) were putatively admixed with wheat genotypes. Such prevalence suggests that the release of transgenic wheat could lead to introgression levels in *Ae. triuncialis* higher than those allowed for cultivar seed stocks (i.e. 0.9% according to EU regulations, Wang *et al.* 2009).

According to our ABC models, the number of immigrating wheat genotypes corresponds to approximately 0.027% of the carrying capacity of the *Ae. triuncialis* metapopulation (Fig. 3). This immigration rate (see below) most likely reflects the high pollen pressures that wheat fields exert on *Ae. triuncialis* (Loureiro *et al.* 2007b). Yet, the reproductive barrier represented by selfing in *Ae. triuncialis* allows only a fraction (i.e. 3%) of these immigrants to effectively reproduce within the resident populations. As a result, the achieved introgression rate of wheat alleles reaches 9.45×10^{-6} [i.e. from $N_{immig} \times (1 - S_{Ae.triuncialis})$, with widest CI = $1.08 \times 10^{-6} – 4.80 \times 10^{-5}$] of the total carrying capacity of *Ae. triuncialis* populations. Although this estimate is lower than spontaneous hybridization rates typically observed at close proximity from fields (ranging between 1.4×10^{-4} and 3.9×10^{-3}, according to David *et al.* 2004; Loureiro *et al.* 2006, 2007a; Arrigo *et al.* 2011) it still shows, using molecular evidence, that an appreciable number of viable and fertile hybrids steadily transfer

wheat DNA to *Ae. triuncialis* every year. To put these figures into perspective, the Grazalema National Park extends over 7,500 Ha of pastures (i.e. the typical environment where *Ae. triuncialis* grows, van Slageren 1994). Conservatively assuming a density of one plant per square meter (this estimate actually ranges from 1 to 650 according to DiTomaso *et al*. 2001), we could expect that ~ 710 hybrids occur every year solely in the investigated area. Although this back-of-the-envelope calculation suffers from large uncertainty (with HPD 95% ranging between 80 and 3,600 hybrids), it indicates that low immigration rates and high selfing might yield many hybrids when appreciated over the whole landscape (but see discussion about post-zygotic barriers below). Our model also shows that introgression leaves appreciable signatures in natural populations (i.e. 10% of the analyzed specimens were admixed) over a short time period. Accordingly, it takes less than 20 generations, after the introduction of wheat cultivars in an area initially free of wheat alleles, to reproduce the empirical admixture signature found in this study. This delay is remarkably short when taken from an evolutionary perspective, especially when considering that it stands mostly for neutral variation (one could expect faster introgression rates for adaptive alleles, see below).

The spatial structure of introgression was more difficult to validate statistically. Several recent studies clearly showed that introgression occurs more frequently in populations bordering cultivated fields than in those growing in remote areas (David *et al.* 2004; Weissmann *et al.* 2005; Zaharieva & Monneveux 2006; Arrigo *et al.* 2011; Parisod *et al.* 2013). The present study partly confirms these results because all the admixed specimens were detected in populations bordering cultivated fields. Accordingly, model 5 yielded the highest posterior probability with regard to the best fit with the empirical signatures of admixture. In addition, estimating model

parameters from population subsets suggested increased wheat immigration rates for populations neighboring the cultivations (see Supplementary Material 2). This adds another line of evidence towards the role played by the spatial context in which admixture occurs, despite estimation of the actual shape of the wheat dispersal kernel proved difficult, and the effect of geographic proximity to cultivated fields could not be strictly validated statistically via model comparison. We attribute these limitations to our sampling effort that might not provide the spatial resolution necessary to reach stronger statistical support (see further discussion about that matter in Supplementary Material 2). In addition to the uncertainty associated with model selection, the large confidence intervals of our parameter estimates might be attributed in part to the dominant nature and genotyping errors inherent to AFLPs. Nevertheless, we are confident that these limitations were unlikely to bias our conclusions, given that genotyping errors would have affected all specimens equally, due to randomized AFLP reactions.

Model assumptions and inference limits

ABC proved more flexible than most available coalescent-based approaches as it allowed to explicitly account for unbalanced spatial sampling in a way to limit estimation biases. However, our models were necessarily based on simplifying assumptions that are discussed here. Most importantly, fitness effects of wheat alleles expressed in the *Ae. triuncialis* genomic background were ignored. Hence, our results hold for neutral markers and different estimates might be obtained with loci affecting fitness in a beneficial (e.g. transgenes conferring resistances to natural pests or pathogens, Perez-Jones *et al.* 2010; Kalinina *et al.* 2011; Kumar *et al.* 2014) or deleterious way (e.g. gametocidal loci, Chase *et al.* 2010). For instance, an

adaptive introgression of beneficial wheat alleles into *Ae. triuncialis* could lead respectively to over- and underestimates for the wheat immigration and the *Ae. triuncialis* selfing rates. This bias is however likely to be mild because most AFLP loci are generally considered neutral (only 2% - 10% of loci screened via genome scans reach differentiation levels detected by FST outlier tests, Bierne *et al.* 2011; Bothwell *et al.* 2013). On the other hand, our models did not account for post-zygotic barriers to gene flow (Feldman *et al.* 2012). Indeed, first generation wheat x *Aegilops* hybrids generally show decreased fertility (Kimber & Feldman 1987), owing to aneuploidy (e.g. meiosis pairing irregularities occur when bread wheat, a hexaploid species, hybridizes with diploid or tetraploid *Aegilops* species), gametocidal genes (Endo & Tsunewaki 1975; Endo 1990) or cyto-nuclear conflicts (Wilson & Driscoll 1983; Ikeda *et al.* 1994). This fitness reduction generally affects pollen production in hybrids, while female fertility subsists and allows the production of viable seeds via backcrossing (Schoenenberger *et al.* 2006). It is therefore possible that post-zygotic barriers are further at work in our study system. We would expect these to further decrease the achieved introgression of wheat alleles into *Ae. triuncialis,* a pattern that our model could have captured by overestimating selfing or underestimating wheat immigration rates. We thus consider that our results are conservative, since the actual introgression rates are likely to be underestimated by our models.

Conclusion

To our knowledge, the present study is among the first to apply process-based models to address crop to wild gene flow. Here, ABC proves to be a powerful tool to explore alternative evolutionary scenarios, and to disentangle and co-estimate the effects of antagonistic evolutionary forces. Accordingly, we estimate the strength of

wheat immigration and that of reproductive barriers, and provide results consistent with empirical evidence of wheat introgression into *Ae. triuncialis*. Our results show that wheat genotypes are steadily migrating within *Ae. triuncialis* populations, probably as a result of strong wheat pollen pressures occurring close to cultivated fields. Moreover, our study shows that selfing can limit, but not prevent, the introgression of crop alleles into wild counterparts, and that even rare introgression events (i.e. 9.45×10^{-6} immigrants per resident, at each generation) leave appreciable admixture signatures when taken at realistic spatial and temporal scales. Hence, it is clear that selfing or allopatry will probably not ensure sufficient reproductive isolation. Careful risk assessment for these species remains important, and transgene containment (e.g. fitness mitigation strategies, Weissmann *et al.* 2008; Kwit *et al.* 2011) should be deployed whenever genetically engineered cultivars are released. Although predicting the consequences of such escapes is beyond the focus of the present work, it should be emphasized that *Aegilops triuncialis* is a widespread and common member of Mediterranean agrosystems (including southern Europe, and northern Africa), which has developed into an aggressive invasive weed in the western USA since its unintentional introduction in California. The (likely) scenario of a transgene escaping into this species could therefore pose a major biosafety and patenting issue, potentially extending over a worldwide scale.

Acknowledgments

We are grateful to the team of the Botanical Garden of the University and City of Neuchâtel for their help with plant cultivation. We thank the Parque Nacional de Grazalema for providing sampling authorizations and giving access to their facilities. Camille Roux and Nicolas Alcala provided comments on earlier manuscript versions. This project was funded by the National Centre of

Competence in Research (NCCR) Plant Survival and the PNR-59, two research programs of the Swiss National Science Foundation. N. Arrigo is funded by an SNSF Ambizione research grant (PZ00P3_148224) and an SNSF professorship (PP00P3_144870) awarded to N. Alvarez. S. Neuenschwander is funded by an SNSF research grant (3100A0-138180) to J. Goudet. The authors are grateful to I. Xenarios, head of the Vital-IT High-Performance Computing Centre (www.vital-it.ch), for allowing us to use powerful computational resources.

Data archiving statement

Raw data for this paper are archived in Dryad (doi:10.5061/dryad.7np24).

References

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

Arnaud J-F, Viard F, Delescluse M, Cuguen J (2003) Evidence for gene flow via seed dispersal from crop to wild relatives in Beta vulgaris (Chenopodiaceae): consequences for the release of genetically modified crop species with weedy lineages. *Proceedings of Biological Sciences / The Royal Society*, **270**, 1565–1571.

Arrigo N, Felber F, Parisod C *et al.* (2010) Origin and expansion of the allotetraploid Aegilops geniculata, a wild relative of wheat. *The New Phytologist*, **187**, 1170–1180.

Arrigo N, Guadagnuolo R, Lappe S *et al.* (2011) Gene flow between wheat and wild relatives: empirical evidence from Aegilops geniculata, Ae. neglecta and Ae. triuncialis. *Evolutionary Applications*, **4**, 685–695.

Arrigo N, Holderegger R, Alvarez N (2012) Automated scoring of AFLPs using RawGeno v 2.0, a free R CRAN library. *Methods in Molecular Biology*, **888**, 155–175.

Arrigo N, Tuszynski JW, Ehrich D, Gerdes T, Alvarez N (2009) Evaluating the impact of scoring parameters on the structure of intra-specific genetic variation using RawGeno, an R package for automating AFLP scoring. *BMC Bioinformatics*, **10**, 1–14.

Austerlitz F, Dick CW, Dutech C *et al.* (2004) Using genetic markers to estimate the pollen dispersal curve.

Molecular Ecology, **13**, 937–954.

Beaumont M A (2010) Approximate Bayesian Computation in Evolution and Ecology. *Annual Review of Ecology, Evolution, and Systematics*, **41**, 379–406.

Beaumont M A, Zhang W, Balding DJ (2002) Approximate Bayesian computation in population genetics. *Genetics*, **162**, 2025–35.

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.

Bothwell H, Bisbing S, Therkildsen NO, Crawford L, Alvarez N, Holderegger R, Manel S (2013) Identifying genetic signatures of selection in a non-model species, alpine gentian (*Gentiana nivalis* L.), using a landscape genetic approach. *Conservation Genetics*, **14**, 467-481.

Chase CD, Ribarits A, Heberle-Bors E (2010) Male Sterility. In: *Plant Developmental Biology-Biotechnology Perspectives* , pp. 437–457. Springer Berlin Heidelberg.

Chen D, Ronald P (1999) A Rapid DNA Minipreparation Method Suitable for AFLP and Other PCR Applications. *Plant Molecular Biology Reporter*, **17**, 53–57.

Csilléry K, Blum MGB, Gaggiotti OE, François O (2010) Approximate Bayesian Computation (ABC) in practice. *Trends in Ecology & Evolution*, **25**, 410–418.

Csilléry K, François O, Blum MGB (2012) abc: an R package for approximate Bayesian computation (ABC). *Methods in Ecology and Evolution*, **3**, 475–479.

D'Andrea L, Felber F, Guadagnuolo R (2008) Hybridization rates between lettuce (Lactuca sativa) and its wild relative (L. serriola) under field conditions. *Environmental Biosafety Research*, **7**, 61–71.

Darmency H, Vigouroux Y, Gestat De Garambé T, Richard-Molard M, Muchembled C (2007) Transgene escape in sugar beet production fields: data from six years farm scale monitoring. *Environmental Biosafety Research*, **6**, 197–206.

David JL, Benavente E, Brès-patry C, Dusautoir J, Echaide M (2004) Biological relevance of polyploidy : ecology to genomics. Are neopolyploids a likely route for a transgene walk to the wild ? The Aegilops ovata x Triticum turgidum durum case. *Biological Journal of the Linnean Society*, **82**, 503–510.

Delplancke M, Alvarez N, Espíndola A *et al.* (2012) Gene flow among wild and domesticated almond species: insights from chloroplast and nuclear markers. *Evolutionary Applications*, **5**, 317–329.

DiTomaso JM, Heise KL, Kyser GB, Merenlender AM, Keiffer RJ (2001) Carefully timed burning can control barb goatgrass. *California Agriculture*, **55**, 47–53.

Eastham K, Sweet J (2002) *Genetically modified organisms (GMOs): The significance of gene flow through pollen transfer*. European Environment Agency, Copenhagen, Denmark.

Ellstrand NC (2012) Over a decade of crop transgenes out-of-place. In: *Regulation of Agricultural Biotechnology: The United States and Canada.* (eds Wozniak CA, McHughen A), pp. 123–135. Springer Netherlands, Dordrecht.

Endo TR (1990) Gametocidal chromosomes and their induction of chromosome mutations in wheat. *The Japanese Journal of Genetics*, **65**, 135–152.

Endo TR, Tsunewaki K (1975) Sterility of Common Wheat With Aegilops triuncialis Cytoplasm. *Journal of Heredity*, **66**, 13–18.

Falush D, Stephens M, Pritchard JK (2007) Inference of population structure using multilocus genotype data: dominant markers and null alleles. *Molecular Ecology Notes*, **7**, 574–578.

Fan X, Sha L-N, Yu S-B *et al.* (2013) Phylogenetic reconstruction and diversification of the Triticeae (Poaceae) based on single-copy nuclear Acc1 and Pgk1 gene data. *Biochemical Systematics and Ecology*, **50**, 346–360.

Felber F, Kozlowski G, Arrigo N, Guadagnuolo R (2007) Genetic and ecological consequences of transgene flow to the wild flora. *Advances in Biochemical Engineering/Biotechnology*, **107**, 173–205.

Feldman M, Levy AA (2009) Genome evolution in allopolyploid wheat--a revolutionary reprogramming followed by gradual changes. *Journal of Genetics and Genomics*, **36**, 511–518.

Feldman M, Levy AA, Fahima T, Korol A (2012) Genomic asymmetry in allopolyploid plants: wheat as a model. *Journal of Experimental Botany*, **63**, 5045–5059.

FitzJohn RG, Armstrong TT, Newstrom-Lloyd LE, Wilton AD, Cochrane M (2007) Hybridisation within Brassica and allied genera: evaluation of potential for transgene escape. *Euphytica*, **158**, 209–230.

Foll M, Beaumont MA, Gaggiotti O (2008) An Approximate Bayesian Computation Approach to Overcome Biases That Arise When Using Amplified Fragment Length Polymorphism Markers to Study Population Structure. *Genetics*, **179**, 927–939.

Gao H, Williamson S, Bustamante CD (2007) A Markov Chain Monte Carlo Approach for Joint Inference of

Population Structure and Inbreeding Rates from Multilocus Genotype Data. *Genetics*, **176**, 1635–1651.

Gompert Z, Lucas LK, Fordyce JA, Forister ML, Nice CC (2010) Secondary contact between Lycaeides idas and L. melissa in the Rocky Mountains: extensive admixture and a patchy hybrid zone. *Molecular Ecology*, **19**, 3171–3192.

Gugerli F, Englisch T, Niklfeld H *et al.* (2008) Relationships among levels of biodiversity and the relevance of intraspecific diversity in conservation – a project synopsis. *Perspectives in Plant Ecology, Evolution and Systematics*, **10**, 259–281.

Hammer K (1980) Vorarbeiten zur monographischen darstellung von wildpflanzensortimenten: Aegilops L. *Die Kulturpflanze*, **28**, 33–180.

Van Heerwaarden J, Ortega Del Vecchyo D, Alvarez-Buylla ER, Bellon MR (2012) New genes in traditional seed systems: diffusion, detectability and persistence of transgenes in a maize metapopulation. *PloS ONE*, **7**, e46123.

Hegde SG, Waines JG (2004) Hybridization and Introgression between Bread Wheat and Wild and Weedy Relatives in North America. *Crop Science*, **44**, 1145–1155.

Ikeda TM, Shibasaka M, Tsunewaki K (1994) Respiratory Activities, Cytochrome Composition and Cytochrome c Oxidase Subunit II Levels of Mitochondria from Alloplasmic Wheats Having Aegilops Cytoplasms. *Plant & Cell Physiology*, **35**, 779–784.

Kalinina O, Zeller SL, Schmid B (2011) Competitive performance of transgenic wheat resistant to powdery mildew. *PloS ONE*, **6**, e28091.

Kilian B, Mammen K, Millet E *et al.* (2011) Aegilops. In: *Wild Crop Relatives: Genomic and Breeding Resources* (ed Kole C), pp. 1–76. Springer Berlin Heidelberg, Berlin, Heidelberg.

Kimber G, Feldman M (1987) Wild wheat. An introduction. In: *Special Report, College of Agriculture* , p. 353. University of Missouri.

Kumar S, Wang Z, Banks TW *et al.* (2014) Lr1-mediated leaf rust resistance pathways of transgenic wheat lines revealed by a gene expression study using the Affymetrix GeneChip® Wheat Genome Array. *Molecular Breeding*, 1–15.

Kwit C, Moon HS, Warwick SI, Stewart CN (2011) Transgene introgression in crop relatives: molecular evidence and mitigation strategies. *Trends in Biotechnology*, **29**, 284–293.

Loureiro I, Escorial MC, García-Baudin JM, Chueca MC (2007a) Hybridization between wheat (Triticum aestivum) and the wild species Aegilops geniculata and A. biuncialis under experimental field conditions. *Agriculture, Ecosystems & Environment*, **120**, 384–390.

Loureiro I, Escorial C, García-Baudin J, Chueca MC (2009) Spontaneous wheat-Aegilpos biuncialis, Ae. geniculata and Ae. triuncialis amphiploid production, a potential way of gene transference. *Spanish Journal of Agricultural Research*, **7**, 614–620.

Loureiro I, Escorial MC, García-Baudín JM, Chueca MC (2006) Evidence of natural hybridization between Aegilops geniculata and wheat under field conditions in Central Spain. *Environmental Biosafety Research*, **5**, 105–109.

Loureiro I, Escorial MC, González-Andujar JL, García-Baudin JM, Chueca MC (2007b) Wheat pollen dispersal under semiarid field conditions: potential outcrossing with Triticum aestivum and Triticum turgidum. *Euphytica*, **156**, 25–37.

Lu B, Snow AA (2005) Gene Flow from Genetically Modified Rice and Its Environmental Consequences. *BioScience*, **55**, 669–678.

Lu B-R, Yang C (2009) Gene flow from genetically modified rice to its wild relatives: Assessing potential ecological consequences. *Biotechnology Advances*, **27**, 1083–1091.

Lynch M (2010) Evolution of the mutation rate. *Trends in Genetics*, **26**, 345–352.

Massinga RA, Al-Khatib K, St. Amand P, Miller JF (2003) Gene flow from imidazolinone-resistant domesticated sunflower to wild relatives. *Weed Science*, **51**, 854–862.

Neuenschwander S, Hospital F, Guillaume F, Goudet J (2008) quantiNemo: an individual-based program to simulate quantitative traits with explicit genetic architecture in a dynamic metapopulation. *Bioinformatics*, **24**, 1552–1553.

Parisod C, Definod C, Sarr A, Arrigo N, Felber F (2013) Genome-specific introgression between wheat and its wild relative Aegilops triuncialis. *Journal of Evolutionary Biology*, **26**, 223–228.

Perez-Jones A, Martins BAB, Mallory-Smith CA (2010) Hybridization in a Commercial Production Field between Imidazolinone-Resistant Winter Wheat and Jointed Goatgrass (Aegilops cylindrica) Results in Pollen-Mediated Gene Flow of Imi1. *Weed Science*, **58**, 395–401.

Schoenenberger N, Guadagnuolo R, Savova-Bianchi D, Küpfer P, Felber F (2006) Molecular Analysis,

Cytogenetics and Fertility of Introgression Lines From Transgenic Wheat to Aegilops cylindrica Host. *Genetics*, **174**, 2061–2070.

Van Slageren MW (1994) Wild Wheats: A Monograph of Aegilops L. and Amblyopyrum (Jaub. & Spach) Eig (Poaceae). *Wageningen Agricultural University Papers*, **94**.

Snow AA (2012) Illegal gene flow from transgenic creeping bentgrass: the saga continues. *Molecular Ecology*, **21**, 4663–4664.

Snow AA, Culley TM, Campbell LG *et al.* (2010) Long-term persistence of crop alleles in weedy populations of wild radish (Raphanus raphanistrum). *The New Phytologist*, **186**, 537–548.

Stewart CN, Halfhill MD, Warwick SI (2003) Transgene introgression from genetically modified crops to their wild relatives. *Nature reviews. Genetics*, **4**, 806–817.

Sunnaker M, Busetto AG, Numminen E *et al.* (2013) Approximate Bayesian Computation. *PLoS Computational Biology*, **9**, e1002803.

Tranel J, Wassom J, Jeschke R, Rayburn L (2002) Transmission of herbicide resistance from a monoecious to a dioecious weedy Amaranthus species. *Theoretical and Applied Genetics.*, **105**, 674–679.

Trucco F, Tatum T, Rayburn AL, Tranel PJ (2009) Out of the swamp: unidirectional hybridization with weedy species may explain the prevalence of Amaranthus tuberculatus as a weed. *The New Phytologist*, **184**, 819–827.

Uwimana B, D'Andrea L, Felber F *et al.* (2012) A Bayesian analysis of gene flow from crops to their wild relatives: cultivated (Lactuca sativa L.) and prickly lettuce (L. serriola L.) and the recent expansion of L. serriola in Europe. *Molecular Ecology*, **21**, 2640–2654.

Waines JG, Hegde SG (2003) Intraspecific Gene Flow in Bread Wheat as Affected by Reproductive Biology and Pollination Ecology of Wheat Flowers. *Crop Science*, **43**, 451–463.

Wang T, Shi Y, Li Y, Darmency H (2009) Testing coexistence and genetic containment for an autogamous crop. *Transgenic Research*, **18**, 809–813.

Warwick SI, Légère A, Simard M-J, James T (2008) Do escaped transgenes persist in nature? The case of an herbicide resistance transgene in a weedy Brassica rapa population. *Molecular Ecology*, **17**, 1387–1395.

Wegier A, Piñeyro-Nelson A, Alarcón J *et al.* (2011) Recent long-distance transgene flow into wild populations conforms to historical patterns of gene flow in cotton (Gossypium hirsutum) at its centre of origin.

Molecular Ecology, **20**, 4182–4194.

Weissmann S, Feldman M, Gressel J (2005) Sequence evidence for sporadic intergeneric DNA introgression from wheat into a wild Aegilops species. *Molecular Biology and Evolution*, **22**, 2055–2062.

Weissmann S, Feldman M, Gressel J (2008) Hypothesis: Transgene establishment in wild relatives of wheat can be prevented by utilizing the Ph1 gene as a senso stricto chaperon to prevent homoeologous recombination. *Plant Science*, **175**, 410–414.

Wilson P, Driscoll C (1983) Hybrid wheat. In: *Heterosis* Monographs on Theoretical and Applied Genetics. (ed Frankel R), pp. 94–123. Springer Berlin Heidelberg, Berlin, Heidelberg.

Zaharieva M, Monneveux P (2006) Spontaneous Hybridization between Bread Wheat (L.) and Its Wild Relatives in Europe. *Crop Science*, **46**, 512–527.

Tables

Table 1. ***Aegilops triuncialis*** **populations included in the study. The geographical coordinates, distance to cultivated fields and the sampling effort of populations are provided along with diversity estimates and F-statistics.**

Population code	NDD[1]	EDD[1]	Distance to cultivated fields[1]	N[2]	PLP[3]	FST[4]	FIS[4]
731227	36.732	-5.228	0 m	23	36%	0.63 (0.54-0.72)	1.00 (0.99-1.00)
753210	36.753	-5.210	0 m	17	35%	0.70 (0.59-0.80)	1.00 (0.99-1.00)
759205	36.760	-5.206	0 m	16	34%	0.70 (0.60-0.80)	1.00 (0.98-1.00)
764152	36.764	-5.152	0 m	9	26%	0.81 (0.58-0.99)	0.99 (0.92-1.00)
777626	36.774	-5.627	0 m	19	34%	0.78 (0.65-0.93)	1.00 (0.96-1.00)
778180	36.778	-5.180	0 m	19	35%	0.87 (0.75-0.96)	1.00 (0.99-1.00)
780518	36.781	-5.519	0 m	18	33%	0.66 (0.56-0.76)	1.00 (0.98-1.00)
784147	36.784	-5.147	0 m	15	32%	0.74 (0.61-0.90)	1.00 (0.99-1.00)
794150	36.795	-5.150	0 m	15	31%	0.63 (0.43-0.88)	1.00 (0.99-1.00)
797223	36.797	-5.223	0 m	17	35%	0.88 (0.80-0.98)	1.00 (0.99-1.00)
797285	36.798	-5.285	0 m	15	38%	0.63 (0.54-0.72)	1.00 (0.99-1.00)
801526	36.801	-5.285	0 m	16	38%	0.63 (0.53-0.74)	1.00 (0.98-1.00)
814500	36.815	-5.501	0 m	18	29%	0.79 (0.69-0.89)	1.00 (0.99-1.00)
818549	36.818	-5.549	0 m	16	32%	0.75 (0.65-0.84)	1.00 (0.98-1.00)
824598	36.825	-5.599	0 m	9	44%	0.88 (0.76-0.99)	1.00 (0.99-1.00)
849587	36.850	-5.587	0 m	15	31%	0.77 (0.64-0.95)	1.00 (0.99-1.00)
702466	36.703	-5.467	100 m	17	37%	0.64 (0.55-0.72)	1.00 (0.98-1.00)
748346	36.749	-5.347	100 m	14	34%	0.82 (0.63-0.95)	1.00 (0.99-1.00)
755237	36.756	-5.237	100 m	9	41%	0.93 (0.84-0.99)	1.00 (0.98-1.00)
757282	36.758	-5.282	100 m	17	38%	0.62 (0.52-0.71)	1.00 (0.98-1.00)
769355	36.770	-5.356	100 m	14	34%	0.76 (0.63-0.89)	1.00 (0.99-1.00)
774273	36.774	-5.274	100 m	17	36%	0.64 (0.55-0.73)	1.00 (0.99-1.00)
785499	36.785	-5.499	100 m	19	33%	0.65 (0.57-0.73)	1.00 (0.98-1.00)
789353	36.790	-5.353	100 m	17	40%	0.63 (0.48-0.78)	1.00 (0.99-1.00)
804302	36.804	-5.302	100 m	17	36%	0.70 (0.58-0.86)	1.00 (0.99-1.00)
715451	36.716	-5.452	1000 m	14	36%	0.83 (0.70-0.96)	1.00 (0.98-1.00)
760490	36.761	-5.490	1000 m	3	---	---	---
762359	36.762	-5.359	1000 m	15	32%	0.67 (0.56-0.80)	1.00 (0.99-1.00)
787325	36.788	-5.326	1000 m	8	31%	0.72 (0.47-0.98)	0.99 (0.90-1.00)
E0613B	36.759	-5.386	1000 m	8	31%	0.50 (0.24-0.92)	0.99 (0.94-1.00)
695422	36.695	-5.422	8500 m	12	29%	0.87 (0.77-0.96)	1.00 (0.97-1.00)

[1]Geographical coordinates (decimal degrees, NDD: latitude, EDD: longitude) and distance to the nearest cultivated field (estimated from aerial views).

[2]Number of analyzed specimens

[3]Within-population diversity, computed as the percentage of polymorphic loci using a rarefaction procedure to account for sampling variation among populations.

[4]F-statistics, estimated as the mode of posterior distributions, 95% highest posterior density intervals provided between brackets.

Table 2. Pairwise comparisons of ABC models assessing the respective importance of secondary dispersal, wheat introgression and spatial structuring of introgression (see Fig. 2 for model definitions). Each comparison is based on posterior probabilities estimated from the 1,000 best simulations, over 200,000 simulations computed for each compared model. The robustness of model selections (reported here as P-values that reflect false-positive rates) was checked using cross-validations (see Supplementary Material 2 for further details). The best fitting models are outlined in bold.

Model parameter	Model comparison	P(model | empirical data)	P-value	Conclusion
Secondary dispersal	**Models 1, 3, 5** versus models 2, 4, 6	**0.98** versus 0.02	0.02*	Exclusion of secondary dispersal.
Interspecific introgression	Model 1 versus **model 3**	0.00 versus **1.00**	0.00***	Presence of interspecific introgression.
Interspecific introgression	Model 1 versus **model 5**	0.00 versus **1.00**	0.00***	
Spatial structure of introgression	Model 3 versus **model 5**	0.32 versus **0.68**	0.33^{NS}	Cannot discriminate between models.

Figure captions

Fig. 1. Geographical locations and wheat admixture levels of *Ae. triuncialis* populations sampled across the Grazalema mountain range (Spain). Briefly, wheat and *Ae. triuncialis* specimens were clustered into $K = 2$ genetic groups using InStruct (Gao *et al*. 2007). This clustering successfully discriminated wheat from *Ae. triuncialis* genotypes, and produced assignment probabilities to the wheat genetic cluster. These probabilities were used as proxies of wheat admixture and are summarized here at the population level using histograms (considering five admixture categories, ranging from Pwheat = 0.00 to 0.36; for simplicity, histogram axes are displayed for only one population). Populations are color-coded according to their geographical proximity to cultivations (white: remote, gray: intermediate distance, black: close proximity). Crosses indicate towns, and gray patches correspond to current crop lands (including wheat fields) in the study area.

Fig. 2. ABC models and parameters. See Materials and methods for detailed explanations. **a)** Introgression scenarios investigated during the model selection procedures (see Table 2 and Supplementary Material 2). The models differ in their presence / absence of secondary dispersal within the *Ae. triuncialis* metapopulation and the mode of wheat introgression (absent, uniform or spatially structured). Parameter values were estimated from the best fitting model using ABC (Figure 3).

Four parameters were explored: the number of biological generations needed to reproduce the empirical introgression patterns (N_{gen}), the wheat immigration rate (N_{immig}, jointly estimated from $M_{wheat\text{-}to\text{-}\textit{Aegilops}}$ - the number of wheat outmigrants and $N_{\textit{Ae. triuncialis}}$ - the population size of *Ae. triuncialis*), the selfing rate of *Ae. triuncialis* ($S_{Ae.\ triuncialis}$) and the spatial dispersal kernel of wheat alleles (*c*). **b)** Metapopulation model, as implemented using forward individual-based simulations (Neuenschwander *et al*. 2008). **c)** Spatial dispersal kernel. The probability of wheat dispersing to an *Ae. triuncialis* population decreases as a function of the population's distance to the wheat field.

Fig. 3. ABC posterior distributions. Parameters were estimated for the two best fitting models (i.e. models 3 and 5, see Fig. 2 and Supplementary Material 2 for details) using custom scripts. Briefly, 1,200,000 simulations were performed for each model, using input parameter values sampled from prior distributions (displayed as light gray curves). The output of each simulation was then compared to empirical observations, using Euclidean distances based on 124 collected summary statistics. We then selected the best 3,000 simulations, in each model, to assess which parameter values yield the most realistic simulations. The obtained parameter values were then corrected using neural-nets implemented in the "abc" R CRAN package and displayed as posterior distributions (**a to d**, the results for models 3 and 5 are displayed in blue and yellow, respectively); with mode and confidence intervals (HPD 95%) indicated. Note that model 3 assumed spatially homogeneous introgression and did not rely on a spatial dispersal kernel.

Figure 1

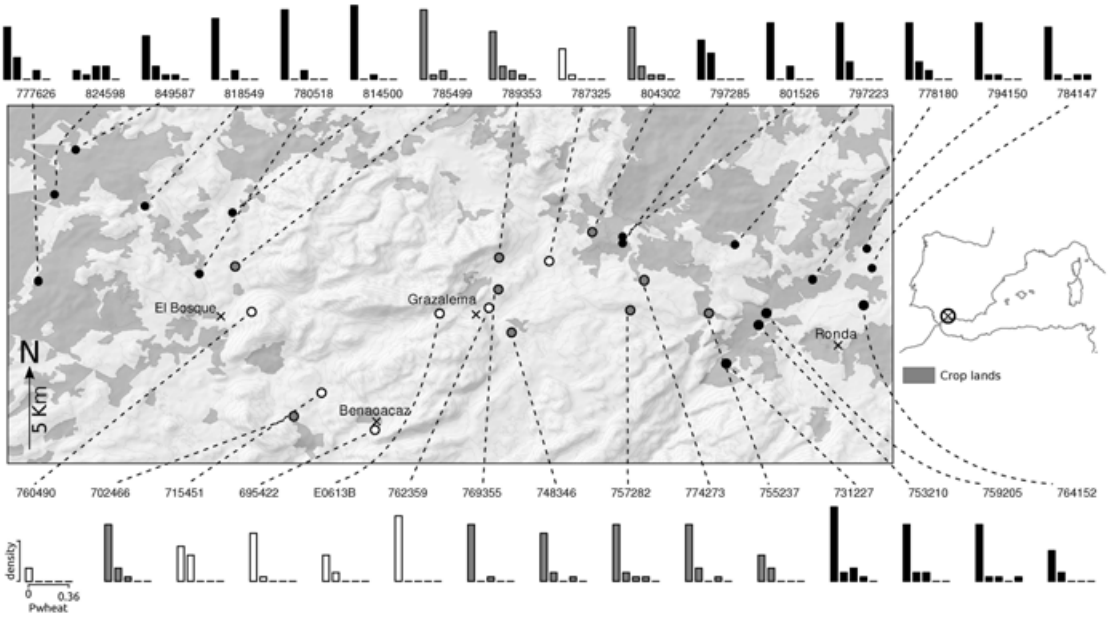

Figure 2

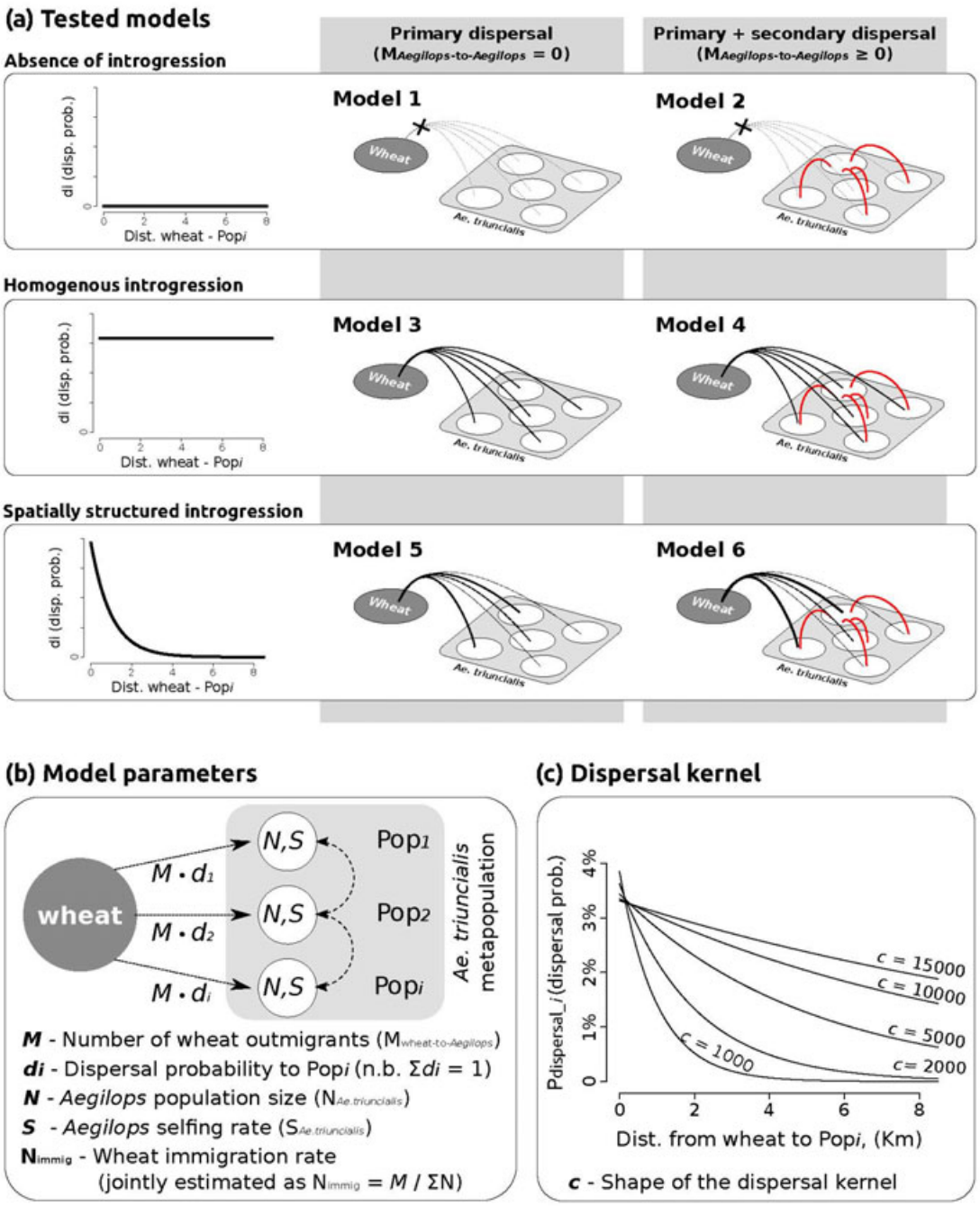

Figure 3