Molecular toolkit and guidelines for the management of genetic pollution

Molecular toolkit and guidelines for the management of genetic pollution

Colophon

Status: Final

Report number: 2018.018

Date of issue: 3-9-2018

Title: Molecular toolkit and guidelines for the management of genetic pollution


Pictures cover: Italian crested newt: Jelger Herder
Pond: RAVON

Pages: 18

Project manager: Ben Wielstra

Funded by: European Union’s Horizon 2020 research and innovation programme: Marie Skłodowska-Curie actions, The Royal Society, Netherlands Organisation for Scientific Research.

© 2018 Stichting RAVON, Nijmegen
Table of contents

Summary

Acknowledgments

1 Invasive species and genetic pollution 1

2 A molecular toolkit to efficiently screen for genetic pollution 2

3 Guidelines for the management of genetic pollution 6
   3.1 When should we intervene? 6
   3.2 Which individuals are eligible for removal? 7
   3.3 What would a management action look like in practice? 8
   3.4 How do we ensure the effectiveness of management actions? 8
   3.5 Decision-tree 9

4 Appendix – Protocol for efficient screening for genetic pollution 10
   4.1 Methods 10
   4.2 Results 11
   4.3 Assessment and recommendations 12

5 References 14
Summary

Genomics plays a crucial role in biological invasion studies. Genetic screening is a key component of management actions concerning invasive hybridizers. We present a molecular toolkit that offers an efficient way to obtain genetic data. With the identification of invasive hybridizing species and their genetically admixed offspring no longer an empirical challenge, the main remaining hurdle in tackling the complex conservation issue of anthropogenic hybridization is establishing clear and implementable policy. We provide guidelines here that we hope help legislators to draft such policy.
Acknowledgments

This report results from a secondment – implemented at the Invasive Alien Species Team of the Office for risk assessment and research of the Netherlands Food and Consumer Product Safety and hosted by Ir. J.W. Lammers – that has been conducted in the context of a Marie Skłodowska-Curie fellowship awarded to Dr. B. Wielstra (with Prof. H.B. Shaffer at the University of California, Los Angeles, USA and Prof. R.K. Butlin at the University of Sheffield, UK acting as hosts). This project has received funding from the European Union’s Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No. 655487. The work was also funded by a Newton International Fellowship awarded to BW. JWA is supported by NWO (ALW Project 824.14.014). Genotyping was conducted at the SNP genotyping line at Leiden University. Reptile, Amphibian and Fish Conservation Netherlands (RAVON) obtained legal consent to conduct tissue sampling (WoD RAVON 12.01).
1 Invasive species and genetic pollution

Increased globalisation enables species to reach locations previously inaccessible to them (1-3). Invasive species pose a major environmental threat: they can endanger populations of native species with extinction through predation, competition and transmission of diseases (4, 5). A more insidious threat is genetic pollution (6-8): the (partial) replacement of local genotypes via hybridization and introgression (9-15). Anthropogenically-induced introgressive hybridization has important implications from a conservation perspective, because it results in genetic replacement and hence a loss of biodiversity at the level of the gene. It also can have far reaching ecological consequences, particularly if hybrids have greater fitness than pure native genotypes (16). Assuming that the goal of conservation biology is to restore native species genetically, as well as the ecological integrity of their communities, both the invader and its genetically admixed offspring would need to be removed. Addressing genetic pollution is a notoriously contentious issue (17-21). With genetic pollution representing a cryptic process and a potential legal minefield, complications arise at the stage of obtaining and interpreting information. In this report our goal is to facilitate the management of genetic pollution by 1) designing a molecular toolkit that allows efficient screening for genetic pollution, and 2) providing guidelines for establishing much-needed policy for the management of genetic pollution.
2 A molecular toolkit to efficiently screen for genetic pollution

Because hybridizing species are generally phenotypically similar, distinguishing pure natives from invasives and particularly from admixed individuals is often extremely difficult based on morphological criteria. The use of genetic data alleviates this problem (21-24). Traditionally, this kind of genetic identification has utilized mitochondrial DNA markers (used for example in “environmental DNA” analyses). Mitochondrial DNA provides a fast and cheap first pass for evaluating the extent of contaminated area (25-27). However, as up to 100% of contaminated individuals in a population may escape detection due to asymmetric introgression (28-32), mitochondrial DNA can only provide a lower estimate of the geographical extent of genetic pollution. Furthermore, mitochondrial DNA does not provide information on the ‘depth’ of genetic pollution, as an individual possessing introgressed invasive mitochondrial DNA might be almost identical to the native species in its nuclear genome.

Nuclear DNA is required to provide a more accurate picture of genetic pollution (33-42). Based on a panel of species-diagnostic single-nucleotide polymorphisms (SNPs), the extent and degree of genetic admixture of a native species and a non-native invasive can be determined, often with great precision (43). Such information is essential in turning the abstract conservation problem of genetic pollution into a set of concrete, actionable decisions.

If an invasive hybridizer is introduced inside the range of a threatened native, this results in an expanding core of heavily contaminated populations, surrounded by an edge of mildly contaminated populations (Box 1). Practical considerations aside, management decisions for core contaminated populations, once identified, would likely be identical to those for “regular” invasive species, with the only difference being that some native genes may be present on a mostly invasive genomic background. Yet, in partially contaminated populations, individuals with either a pure native genotype or individuals expressing a low level of genetic pollution are present. Total elimination of such populations might be undesirable for conservation reasons if the native species is threatened and/or legal reasons if the native species is protected. In such instances, the individuals that qualify for removal or protection might be identified based on their genetic composition (21).

To distinguish 1) core contaminated, partially contaminated, and unaffected populations, and 2) individuals eligible for elimination or protection within partially contaminated populations, a management action would entail catching a large number (thousands) of individuals and holding them in captivity until their genotype has been established and their fate (remove or release) determined. Such a plan requires reliable genotyping and a fast turnaround measured in hours-to-days (not weeks) and needs to be financially feasible. Although consulting more markers would result in a higher accuracy (Box 2), the time and costs involved would increase as well. In practice we need to be able to genotype a moderate number of SNPs for a large number of individuals, in a short time span and at reasonable costs. In the Appendix we outline a protocol employing the Kompetitive Allele Specific PCR (KASP) genotyping system (LGC genomics, UK) (44). We provide proof of concept using a case study on crested newts (Box 3).
Box 1 - Conceptualization of genetic pollution

After an invasive hybridizing species is introduced inside the range of a native species (Fig. 1a), it expands at the expense of the native species and the hybrid zone between the species moves outwards from the introduction site (Fig. 1b). There will be a core of populations with a high frequency of invasive genes (black), while populations currently beyond the reach of genetic pollution will be purely the native species (white). In the area in between, populations with different degrees of genetic admixture occur, with the frequency of invasive alleles decreasing at the outer margins. Particularly in the outer populations, individuals of conservation concern (with either a pure native genotype or levels of genetic pollution deemed sufficiently low) are predicted to be present.

Figure 1 Schematic representation of invasive hybridization. In panel a, an invasive hybridizing species is introduced (black) inside the range of a native species (white). In panel b, the hybrid zone between the species has moved outwards; pure invaders are left in its wake (black), in the hybrid zone itself natives, invaders and genetically admixed individuals are present (gradient) and beyond the hybrid zone only pure natives occur (white).

Box 2 - Data simulation

Using HybridLab (45) we illustrate the detectability of alien alleles under an extreme scenario of repeated backcrossing to a native species with different numbers of diagnostic nuclear markers. We model 1000 offspring each for ten generations of backcrossing with an increasing number of markers (Fig. 2). This exercise shows a positive correlation between the number of markers and the number of generations of backcrossing over which the presence of alien alleles can be detected. Such data simulation can be used to guide management decisions, for example by determining the chance of detecting admixed individuals with different numbers of markers, under particular admixture scenarios.

Figure 2 Detecting genes of an invasive hybridizer under a scenario of repeated backcrossing to a native species. After an initial hybridization event individuals are heterozygous native/invasive for all markers (F1). Over ten generations of backcrossing to the native species (B1-10) the proportion and detectability of invasive alleles continually decreases, but the ability to detect at least some non-native ancestry decreases more slowly as more markers are studied. The dotted line indicates the 95% probability level for detection of at least one alien allele in a sampled individual with random mating and no fitness differences between native and non-native alleles.
Box 3 – Dutch crested newt case

The northern crested newt *T. cristatus* is a European wildlife icon. However, the Italian crested newt *T. carnifex* has been introduced in several places inside the range of *T. cristatus* and invasive hybridization has been documented from Geneva, France/Switzerland (46, 47), Munich, Germany (48), Surrey, England (49, 50) and the Veluwe, Netherlands (51, 52) (Fig. 3). We present a case study from the Veluwe in the Netherlands for which we possess a tissue bank and background knowledge (51-62). The Veluwe area is positioned well within the native range of *T. cristatus* and *T. carnifex* has been introduced there locally (Fig. 3), presumably in the late 1970s, near the village of Vaassen. The invasive species is expanding at the expense of the native one and the two species hybridize where they meet. The native is difficult to distinguish morphologically from the invader and genetically admixed offspring are even harder to identify using phenotypic traits. The Netherlands has an (inter)national responsibility to protect the native *T. cristatus*: the species is listed as vulnerable on the Dutch Red List (63) and on Appendix II and IV of the EU Habitats Directive and Appendix 2 of the Bern Convention. By genotyping ten diagnostic nuclear DNA markers, we provide tangible results that facilitate the management of the Dutch *Triturus* case. From the core of the introduction site outwards populations (here defined as breeding ponds) range from pure invasive (*T. carnifex*), via decreasing degrees of genetic pollution, to pure native (*T. cristatus*) (Fig. 4). We observe traces of the native species in the populations now genetically almost purely belonging to the invasive species, in line with the expectation that the invasive hybridizer displaced the native species.

![Figure 3](image)

**Table 1** Details on the localities in Figs. 3-4. Per locality twelve individuals were genotyped.

<table>
<thead>
<tr>
<th>No.</th>
<th>Locality</th>
<th>Latitude</th>
<th>Longitude</th>
<th>No.</th>
<th>Locality</th>
<th>Latitude</th>
<th>Longitude</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Hoog Soorensche Bosven</td>
<td>52.242</td>
<td>5.879</td>
<td>7</td>
<td>Leemkuilen Staverden 5</td>
<td>52.269</td>
<td>5.753</td>
</tr>
<tr>
<td>2</td>
<td>Ruiterpont</td>
<td>52.239</td>
<td>5.876</td>
<td>8</td>
<td>Leemkuilen Staverden 2</td>
<td>52.275</td>
<td>5.732</td>
</tr>
<tr>
<td>3</td>
<td>Prinsenkuil</td>
<td>52.298</td>
<td>5.865</td>
<td>9</td>
<td>Leemkuilen Staverden 1</td>
<td>52.275</td>
<td>5.732</td>
</tr>
<tr>
<td>4</td>
<td>Noorderheide 1</td>
<td>52.305</td>
<td>5.819</td>
<td>10</td>
<td>Mostedvoorn Andromedaven</td>
<td>52.349</td>
<td>5.822</td>
</tr>
<tr>
<td>5</td>
<td>Groot Zeilmeer</td>
<td>52.240</td>
<td>5.788</td>
<td>11</td>
<td>Leemkuilen Tongeren</td>
<td>52.362</td>
<td>5.928</td>
</tr>
<tr>
<td>6</td>
<td>Klein Zeilmeer</td>
<td>52.235</td>
<td>5.787</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Box 3 – continued

Individuals from populations 1-3 have a pure or almost completely invasive genotype, while population 11 appears unaffected by genetic pollution (Fig. 4). For populations 4-7 & 9 we find genetic admixture between the native and the invasive species in all surveyed individuals – such populations might be considered eligible for removal in their entirety. In populations 8 and 10, individuals that are unambiguously genetically admixed are found in syntopy with individuals for which no alien alleles were observed – such populations might warrant a survey with a larger number of markers to single out individuals of conservation concern. While populations 1-10 show genetic admixture based on nuclear DNA (Fig. 4), we did not observe non-native mitochondrial DNA in populations 6 and 10 (Fig. 3), suggesting the latter marker underestimates genetic pollution. In fact, about half (62/116, 53%) of individuals identified as possessing alien alleles possessed the native mitochondrial DNA genotype.

Figure 4 Genetic composition of crested newt populations from the Veluwe in the Netherlands based on ten diagnostic nuclear SNPs. The x-axis shows ancestry (the fraction of alleles derived from each parental species) and the y-axis heterozygosity (the fraction of loci heterozygous for alleles from each parental species). The lower left corner of each triangle represents a pure native genotype, the lower right corner a pure invasive genotype, and the upper corner a first generation (F1) hybrid. Results for populations 1-3 are combined and results for populations 4-11 are shown in individual graphs. Note that individuals with identical genotypes overlap. Population numbers correspond to Fig. 3 and Table 1.
3 Guidelines for the management of genetic pollution

The threat to biodiversity posed by genetic pollution, while previously underestimated by conservation biologists (9), is gaining traction in the recent conservation literature (17-21). The main outstanding issue in the management of genetic pollution concerns the translation of a quantification of genetic pollution into concrete management actions. We reiterate that policy on genetic pollution is urgently needed (17-21) and here outline what we believe is a step in this direction.

As with exotic species in general, preventing the introduction of potential invasive hybridizers in the first place is crucial (18, 64). Yet, absolute prevention of the introduction of invasive species is unlikely to succeed and many instances of genetic pollution are already in place. We here focus on the decision-making process in a situation where genetic pollution is ongoing.

Tackling the complex management of genetic admixture between native and invasive species requires interactions between scientists, legislators and conservationists and, ideally, that all parties involved have a basic understanding of genetic results (16). Based on previous experience, we recommend that additional partners, in particular land managers, should also be invited to participate in the early stages of the decision process (65). Finally we emphasize the critical need for public support – which is not straightforward in the field of invasion biology (66-71).

3.1 When should we intervene?

When considering a management action to counter genetic pollution two relevant issues (21) are 1) whether the invasive hybridizer poses a threat of genetic pollution to genetically pure native populations and 2) whether genetically pure native populations are rare. There are precedents, where genetic pollution was deemed a risk while unaffected populations of the native species were still wide-spread, and culling of the invasive species and genetically admixed individuals has been conducted (72-74). However, the rarer native populations unaffected by genetic pollution are, the higher the conservation and restoration value of hybrid populations. Where genetic pollution has progressed to the point that most if not all native populations have been affected, its removal is no longer feasible (21, 75). Under such exceptional conditions, trying to restore the ecological authenticity of the original species has been argued the best course of action (21, 36).

We would argue that invasive species that are genetically similar enough to a native species that the two are capable of hybridizing are typically ecologically similar as well. Hence the two are likely to interact and in consequence the invader is likely to pose a threat to the native by 1) direct competition, 2) wasted reproductive effort, and, if hybrids are capable of reproducing themselves, 3) genetic pollution. The conservation value of hybrids is more difficult to rationalize a priori, as it depends on the size of the range of the native species and the expansion rate of the invasive hybridizer (and perhaps the number of independent introductions that have occurred) and hence will need to be determined on a case-by-case basis. Yet, given that most invasive hybridizers have been introduced on the scale of decades and at most centuries ago, we consider it more likely than not that, on continents
(but perhaps not on oceanic islands) genetic pollution has not yet affected the majority of the range of the native species and hence hybrid populations will generally have a low conservation value.

If invasive hybridizers are considered a threat to native biodiversity and are of low conservation value, then eradicating them could be regarded as an ethical conservation strategy (20). Given the above, we would argue that, in principle, we should aim to negate genetic pollution. However, considering that resources for conservation are limited, in practice prioritization is a likely necessity (64, 76). Impact of the invasive hybridizer and practicality of a potential management action play a key role in this process. The urgency to intervene increases when the native species is of higher conservation concern (17) or when the ecology of invasive hybridizers and genetically admixed offspring differs from the native species (77). The feasibility of a management action increases when the associated impact on community members is low and spatial isolation of polluted populations is attainable (18, 20). In short, an informed decision needs to be taken on whether to intervene or not on a case-by-case basis – and properly assessing the situation most likely requires some basic genetic data. As with invasive species management in general, the earlier a potential management action is started, the higher the chance it will be successful (and the lower the associated costs and impact are) (69). From here on we assume a situation where the decision of negating genetic pollution has been reached.

3.2 **Which individuals are eligible for removal?**

The legal status of an individual that is genetically admixed between a threatened native and a hybridizing invasive is currently unclear in most countries (17, 20) – including the Netherlands. While hybrid-derived individuals whose ancestry includes an endangered species might be legally protected, such individuals would need to be removed in a management action. Hence, it is imperative that a legal route to remove invasive hybridizers and their descendants is available.

Individuals with non-native ancestry below a certain threshold might be considered to have conservation value (21). Each case of genetic pollution may require different genetic thresholds to achieve a specified set of management outcomes, meaning flexibility in decision making is essential (17, 20, 21). The threshold applied for marking a population or individual as ‘worthy to protect’ versus ‘to be eliminated’ is eventually a juridical decision (16).

When a threshold is decided upon, populations and individuals can be identified as genetically polluted. Populations that contain only polluted individuals beyond reasonable doubt could be targeted for direct removal. In the case of populations that also contain individuals of conservation value, there would be two options: 1) complete removal while accepting casualties, or 2) filtering out those individuals of conservation value. While the first option is much more pragmatic, it might constitute a more difficult or impossible legal route, is probably less socially acceptable and would still involve continued action over multiple generations.
A power analysis with the ConGRESS Planning Tool on Hybridization (78, 79) (http://www.congressgenetics.eu) can help to determine the number of markers required to reach a particular threshold with a certain probability under definable assumptions about breeding probabilities (e.g. all individuals mate randomly with respect to genotype) and survival likelihoods (e.g. that native and alien alleles are neutral with respect to fitness).

3.3 What would a management action look like in practice?

Genetic data are pertinent in deciding which populations are to be protected, are eligible for elimination and, in a more complicated scenario, are partially contaminated and need individuals of conservation value to be filtered out (9). After the introduction of an invasive hybridizing species, introgression is (fortunately) expected to be biased towards the expanding invasive species (30, 31). Hence, with a relatively low number of markers and individuals, genetic pollution should be detectable throughout most affected populations. This strategy will allow us to home in on the outer boundary of genetic pollution.

Yet, at the advancing frontier of invasion, repeated backcrossing to the native species would cause invasive DNA to become diluted (Box 1 & 2). Furthermore, even under the condition that an expanding hybrid zone may stabilize (80-83), genetic pollution remains a concern, as adaptive introgression could result in a small fraction of invasive genomes to advance ahead of the overall hybrid zone between the native and invasive species (84, 85). Near the outer edge of genetic pollution more statistical power, achievable by surveying a greater number of markers for many individuals, would be required to ensure that apparently pure native populations are indeed likely not to possess alien alleles. Furthermore, in those populations showing mild genetic pollution, individuals of conservation value can be separated from individuals eligible for removal if needed.

Given the above, we suggest a two-step approach, with 1) an initial scan using a small number of nuclear markers (about ten) to identify completely admixed, partially admixed and apparently unpolluted populations (sampling about ten individuals per population), and 2) a subsequent higher resolution screening, employing a larger number of markers (about a hundred), to confirm apparently clean populations that neighbor polluted populations (sampling at least 50 individuals), and to filter out individuals of conservation concern for mildly admixed populations (sampling all individuals).

3.4 How do we ensure the effectiveness of management actions?

Populations for which genetic pollution has been controlled in a management action must be monitored, using genetic screening, to ensure that only native genotypes remain in and/or re-colonize the population. We recommend collecting samples annually, and analyzing them every 3-5 years. Populations where the introduced hybridizer and genetically admixed individuals have been removed could be re-stocked with genetically pure native individuals from neighbouring populations (i.e. from populations that have been genetically screened and confirmed not to possess alien alleles). In the first place this helps the displaced native species to recover its former range more efficiently, which is particularly relevant if its dispersal distance is low. Furthermore, it impedes the invasive hybridizer to re-colonize controlled populations as a high density native population is
harder to replace than a low density one, a phenomenon referred to as high density blocking (86). In particular cases it has been shown that man-made habitat alteration favors foreign genotypes, and in such cases restoration of the original habitat is likely to provide the native genotype with a competitive edge (17).

3.5 Decision-tree

The framework below summarizes the decision process outlined above.
Appendix - Protocol for efficient screening for genetic pollution

4.1 Methods

Genomic data representing both the native and the invasive species could be used directly for de novo SNP discovery, but an ascertainment bias might be introduced if these data did not include multiple individuals from across the range of both parent species (43, 87, 88). However, obtaining genome-scale data for many individuals is often not feasible, either technically or economically. A workaround is to identify potential variable sequence regions from limited genomic data, sequence these particular regions for a larger set of individuals, and use the resulting dataset for SNP discovery (89).

Based on a single Triturus transcriptome, Wielstra et al. (89) sequenced 52 short (ca 140 bp) nuclear markers positioned in 3' UTR regions of protein-coding genes, and used an Ion Torrent next-generation sequencing protocol to obtain sequence data for three individuals from four populations across the range of both T. carnifex and T. cristatus (Box 3). We used these sequence data (available from Dryad Digital Repository entry http://dx.doi.org/10.5061/dryad.36775) to identify SNPs that occur in two variants, one diagnostic for the native and the other for the invasive species. We focussed on a subset of 15 nuclear markers for which T. carnifex and T. cristatus had species diagnostic allele variants – we expected ten markers would provide enough resolution for our purposes but took into account the risk that primer design or performance for part of the initial 15 markers would probably fail. Additionally, Wielstra et al. (90) used Sanger sequencing to obtain sequence data (658 bp) for one mitochondrial DNA marker from a cross the range of T. carnifex and T. cristatus and we included this marker for SNP discovery. We determined diagnostic SNPs by checking the sequence alignments by eye in MacClade 4.08 (91). Wielstra et al. (52) provide further details.

DNA was extracted with the Qiagen Dneasy 96 Blood & Tissue Kit and the resulting DNA extract was used at a 1:49 dilution. We genotyped 156 crested newts in total to test our protocol (Box 3). We sampled the 12 individuals of both parental species on which marker design was based as well as the 132 individuals studied by Meilink et al. (51) that derive from 11 populations from the Veluwe in the Netherlands (12 individuals per population) and range from pure native (T. cristatus), via different degrees of genetic admixture, to pure invasive (T. carnifex). Wielstra et al. (52) provide further details on sampling.

Genotyping was conducted at the SNP genotyping facility of the Institute of Biology, Leiden University, using the Kompetitive Allele-Specific PCR (KASP) genotyping system (LGC genomics, UK). The KASP technology encompasses fluorescence-based genotyping (44). The SNP variant present in each individual (both variants in the case of a heterozygote) is determined in uniplex assays, based on two allele-specific primers with a final base complementary to one of the two potential SNPs, that also possess a unique tail sequence. Different fluorescently labelled primers present in the KASP master mix correspond to each tail sequence and are activated when incorporated during subsequent PCR cycles, with further cycling causing signal intensity to increase. No library preparation, DNA sequencing and bioinformatics is involved. The KASP technology is economical and accurate compared to other SNP genotyping technologies (44). We designed primers using
the Kraken software (LGC genomics, UK) and ordered them from Integrated DNA Technologies.

The PCR mix was made according to the manufacturer’s instructions. The LGC genomics KASP mix 4.0 was used and PCRs were carried out in 1536 wells plates with a reaction volume of 1 ul in a hydrocycler. The PCR consisted of an initial step of 15 min at 94 ºC, followed by 10 cycles of 20s at 94 ºC and 60s at 61 ºC, followed by 26 cycles of 20s at 94 C and 60s at 55 ºC. After these 36 cycles fluorescence was measured on a Pherastar Plate reader. PCR was continued and after 39, 42 and 45 cycles fluorescence was measured again to follow the trajectory of all samples. Genotypes were automatically called using the Kraken software, visually inspected and occasionally manually corrected.

The nuclear markers in which the targeted SNPs were positioned were previously sequenced using Ion Torrent next-generation sequencing for the same individuals (51) (available from Dryad Digital Repository entry http://dx.doi.org/10.5061/dryad.1f75). We compared the results of the two genotyping strategies for each marker–individual combination to identify potential differences between the KASP genotyping and Ion Torrent protocols. Similarly, mitochondrial DNA sequences were previously obtained by Sanger Sequencing (51) (Genbank accession numbers: Tcri01 = GU982383; Tcri42 = KJ569095; Tcri43 = KJ569096, Tcar03 = JQ653329) and compared with the KASP genotyping results.

4.2 Results

For the mitochondrial marker an assay was successfully designed and all individuals could be genotyped (Box 3). Allocation to species mitochondrial DNA type was identical to results previously obtained by Sanger sequencing. For all 15 candidate nuclear markers, assays could be designed. Genotypes for two of these markers (amot and ibtk) could not be unambiguously called. A further two nuclear SNP markers (cnppd and col18) showed a single instance of heterozygosity in a parental individual, suggesting either a genotyping error or that the markers were not fully diagnostic. Finally, one marker (taf8) was dropped because it showed non-Mendelian inheritance. These five markers were excluded from further analysis. Wielstra et al. (52) provide further details on genotyping results.

For the ten remaining nuclear markers, 1472 out of 1560 (94.4%) attempted SNP calls provided unambiguous results. For the remaining 88 attempted SNP calls, in three instances the PCR failed, in 31 instances no distinction could be made with confidence between heterozygotes and homozygous natives (i.e. the presence of 0 or 1 alien alleles) and in 54 instances we could not distinguish between heterozygous and homozygous invasives (1 or 2 alien alleles). Hence, 54 + 1472 = 1526 (97.8%) of attempted SNP calls provided reliable information on the presence of alien alleles. Ancestry and heterozygosity for the 11 Veluwe populations, based on the KASP genotyping of the nuclear SNPs (excluding SNP calls in which homozygotes and heterozygotes could not be distinguished), were visualized with the R package HIest (92) (Box 3).

For the Ion Torrent protocol for 1497 out of 1560 (96.0%) marker-individual combinations two native and/or invasive alleles could be genotyped unambiguously. Ten marker-individual combinations had no data. An issue particular to the Ion Torrent protocol was that new alleles (not previously identified in the parental species) were found and hence
could not be allocated to species. For ten marker-individual combinations neither allele could be allocated to species, for 21 only one native allele, and for 22 only one alien allele could be identified. Hence, for $1497 + 22 = 1519$ (97.4%) marker-individual combinations we obtained reliable information on the presence of alien alleles.

Genotyping efforts with the KASP genotyping and Ion Torrent protocol provided different calls for 46 out of 1560 (2.9%) genotyping efforts. For 17 incongruent genotyping efforts there was disagreement whether zero or one alien alleles were present. For the remaining 29 genotyping efforts both protocols agreed upon the presence of alien alleles, but disagreed on whether one or two were present. Agreement between the two datasets on the presence of alien alleles using Cohen’s kappa (the number of genotypes consistently called as with or without alien alleles by both protocols, divided by the total number of genotypes called with both protocols), ranged from 0.97 to 1 for individual markers, and for the total dataset Cohen’s kappa was 0.99. The differences between genotype methods occurred in 12 out of 156 (7.7%) individuals, for which Cohen’s kappa ranges from 0.7-0.9. For all these individuals the presence of alien alleles was established unambiguously for some, but not all of the markers. At the level of the total dataset, the KASP genotyping and Ion Torrent protocols agreed that for 128 individuals alien alleles were present whereas for 27 individuals neither protocol could confirm the presence of alien alleles; for one individual the presence of a single alien allele identified with the Ion Torrent protocol could not be confidently established with the KASP genotyping protocol.

4.3 **Assessment and recommendations**

A genotyping method aiming to identify alleles derived from an invasive species should minimize both false negatives (missing alien alleles when they are really present) and false positives (claiming an allele is non-native when it is actually native). Failing to identify and remove a proportion of the genetically admixed individuals, or inadvertently removing pure natives, would both compromise the management goal of restoring native genetic integrity. However, effective management may also require rapid genetic testing and swift decisions, which may come with high cost or error rates.

We are here in a position to compare the KASP genotyping results with a previously generated dataset. The KASP genotyping strategy performed well compared to a more expensive and time-consuming Ion Torrent next-generation sequencing pipeline (51, 89). In spite of minor differences between the two protocols, at the level of the total dataset agreement is essentially perfect and only a single, slightly admixed individual yielded equivocal results. We consider KASP genotyping a reliable and, when a large number of individuals has to be genotyped for a moderate number of SNPs, preferable tool for the detection of genetic admixture between native and invasive species.

Compared to the Ion Torrent and other next-generation protocols the KASP genotyping protocol has some major advantages (44). KASP genotyping is: (i) fast (at the SNP genotyping facility of the Institute of Biology, Leiden University, 100,000 data points can be genotyped in a single day, so for ten markers 10,000 individuals can be genotyped), (ii) relatively affordable if a small number of SNPs is sufficient (at about 0.35 € per SNP when 384 individuals are genotyped at a time, a study with ten markers amounts to 3.5 € per individual, excluding DNA extraction), (iii) simple (PCRs are conducted per individual per marker using an automated pipeline and genotypes are automatically called and available
in a user-friendly format), and (iv) scalable (SNP calls can be repeated or extra SNPs can easily be genotyped at a later stage).

A complication we encountered with KASP genotyping in some of the attempted SNP calls was that it was not always possible to distinguish heterozygotes from homozygotes. Although polyploidy and paralogy could underlie this pattern we consider this unlikely because: (i) polyploids are rare in salamanders and have virtually never been recorded in *Triturus* newts (93-95), (ii) the initial marker design phase excluded multicopy genes (89), and (iii) the distribution of unclear SNP calls was not linked to any particular individuals or markers. An alternative explanation is that ambiguous SNP calls are a consequence of the large genome size of our study system (c. 20-30 Gb in the genus *Triturus* (96)). The larger the genome, the lower the concentration of the targeted SNPs per unit of DNA, which may compromise SNP calling. Another project on *Triturus* using DNA extract diluted 1:9 rather than 1:49 yielded more consistently interpretable results (97). This suggest that our starting amount of DNA was on the low side and we predict elevating the starting DNA levels would further increase genotyping performance.

The rate-limiting step in our protocol is DNA extraction. If there is no requirement to keep extracted DNA after an individual has been genotyped, a fast and cheap but destructive Chelex extraction could be used (including labor and plastics about 0.50 € per individual, a c. twenty-fold decrease in costs compared to the extraction method used in the present study; we estimate c. 1000 individuals could be extracted in a day). Another *Triturus* project using the Chelex method yields high quality genotyping results with the KASP platform (98). Please note that costs concern lab work only and e.g. field work is not included in our calculations.

The molecular toolkit introduced here is directly applicable to other *T. carnifex* introductions that have occurred in the range of *T. cristatus* (Box 3). It can easily be adapted to other cases of invasive hybridization. However, please note that the SNP discovery step needs to be repeated when focussing on new taxa and, if there are no publicly available genetic resources available, would require an initial investment to obtain such resources. Fortunately such genomic resources are now easy to generate (99, 100).
References

42. C. Vilà et al., Combined use of maternal, paternal and bi-parental genetic markers for the identification of wolf–dog hybrids. Heredity 90, 17 (2003).


