Current Zoology 61 (3): 505–519, 2015

Is there evidence of selection in the dopamine receptor *D4* gene in Australian invasive starling populations?

Lee Ann ROLLINS[1,2*], Michael R. WHITEHEAD[1,7], Andrew P. WOOLNOUGH[3,4], Ron SINCLAIR[5], William B. SHERWIN[1,6]

[1] Evolution & Ecology Research Centre, Biological, Earth and Environmental Sciences, University of New South Wales, Sydney, New South Wales, 2052, Australia

[2] Centre for Integrative Ecology, Life and Environmental Sciences, Deakin University, Locked Bag 20000, Geelong, Victoria, 3220, Australia

[3] Biosecurity Division, Department of Economic Development, Jobs, Transport and Resources, 1 Spring Street, Melbourne, Victoria, 3000, Australia

[4] Vertebrate Pest Research Section, Department of Agriculture and Food, 100 Bougainvillea Avenue, Western Australia, 6058, Australia

[5] Biosecurity SA, GPO Box 1671, Adelaide, South Australia, 5001, Australia

[6] Murdoch University Cetacean Research Unit, Murdoch University, South Road, Murdoch, Western Australia 6150, Australia

[7] Current address: Evolution, Ecology and Genetics, Research School of Biology, The Australian National University, Canberra, Australia

Abstract Although population genetic theory is largely based on the premise that loci under study are selectively neutral, it has been acknowledged that the study of DNA sequence data under the influence of selection can be useful. In some circumstances, these loci show increased population differentiation and gene diversity. Highly polymorphic loci may be especially useful when studying populations having low levels of diversity overall, such as is often the case with threatened or newly established invasive populations. Using common starlings *Sturnus vulgaris* sampled from invasive Australian populations, we investigated sequence data of the dopamine receptor D4 gene (*DRD4*), a locus suspected to be under selection for novelty-seeking behaviour in a range of taxa including humans and passerine birds. We hypothesised that such behaviour may be advantageous when species encounter novel environments, such as during invasion. In addition to analyses to detect the presence of selection, we also estimated population differentiation and gene diversity using *DRD4* data and compared these estimates to those from microsatellite and mitochondrial DNA sequence data, using the same individuals. We found little evidence for selection on *DRD4* in starlings. However, we did find elevated levels of within-population gene diversity when compared to microsatellites and mitochondrial DNA sequence, as well as a greater degree of population differentiation. We suggest that sequence data from putatively non-neutral loci are a useful addition to studies of invasive populations, where low genetic variability is expected [*Current Zoology* 61 (3): 505–519, 2015].

Keywords *DRD4*, Starling, Selection, Novelty-seeking behaviour

All loci are affected equally by demographic processes, which determine the strength of genetic drift, but the effects of selection will vary across genes (Cavalli-Sforza 1966; Bowcock et al., 1991) potentially affecting variation within and between populations. As a result, a higher variance in measures of population differentiation is expected when calculated using genes under selection than those that are neutral (Lewontin and Krakauer, 1973). Therefore, Allendorf et al. (2008) have suggested that population data from genes under strong selection may be useful for detecting between-population variation where neutral markers have failed to do so. However, there are caveats for the use of markers under selection. First, the identification of genes that are under selection is problematic, since DNA that is not protein-coding may sometimes be under selection (Pimpinelli and Dimitri, 1989; Andolfatto, 2005). Second, there is a range of possible effects of selection upon gene diversity within-populations, genetic differentiation between populations, and the interaction between them. Depending upon the mode of selection and the similarity of selective pressures across environments, population differentiation in loci under selection may be significantly lower or significantly higher than expected

Received Jan.9, 2015; accepted Apr. 2, 2015.

∗ Corresponding author. E-mail: lee.rollins@deakin.edu.au

from neutral markers. Karl and Avise (1992) demonstrated lowered population differentiation in their comparison of geographic changes to allele proportions of markers under balancing selection (allozymes) to those from putatively neutral mtDNA and anonymous nuclear loci. Conversely, compared to neutral loci, those under different selection regimes in different environments may show enhanced differentiation. This approach has been used to identify genes that are putatively under selection, using genome-wide single nucleotide polymorphism (SNP) scans (Akey et al., 2002) and many studies have demonstrated that F_{ST} values from genes believed to be under strong selection were significantly higher than those estimated from neutral microsatellites (Pampoulie et al., 2006; Hemmer-Hansen et al., 2007). Finally, selection also affects within-population diversity: directional selection may depress or eliminate genetic diversity within populations, whereas balancing selection maintains within-population diversity (Nielsen, 2005).

For the estimation and comparison of within-population gene diversity, functional genes (or linked loci such as introns) may provide different information to microsatellite or mitochondrial sequence data. For example, in a study of eight populations of four different carnivore species, Väli et al. (2008) found that the variance in within-population genetic diversity across populations was approximately ten times greater at nuclear intron loci than at microsatellite loci analysed for the same group of individuals. They concluded that estimates of within-population genetic diversity based on microsatellite data may not be representative of genomic levels of diversity due to ascertainment bias (i.e. the highly polymorphic microsatellites selected for population studies inflate estimates of within-population diversity, thereby reducing variation in diversity across populations). Villablanca et al. (1998) found approximately 30% more genetic diversity at nuclear introns than mitochondrial DNA (mtDNA) in native range medfly populations *Ceratitis capitata*, but when introduced populations were analysed, within-population genetic diversity in introns was approximately three times higher than that found for mtDNA. They concluded that higher levels of genetic diversity found in intron sequence data may render them more useful than other types of markers for understanding invasion history and detecting variation between populations. This is especially true in situations where neutral diversity is expected to be low, such as founder events initiating invasive populations. However, it is important to note that the presence of higher levels of within-population genetic diversity may not, by itself, result in more informative data regarding population history and differentiation between populations.

Although the study of functionally important loci may provide advantages over other types of markers as described above, it is important to consider the effects of selection on demographic and evolutionary history of the gene under study (Zhang and Hewitt, 2003). In some cases of strong selection, population history remains unobscured (e.g. Pampoulie et al., 2006), but other examples show that population history can be completely obscured by strong selection (e.g. Hemmer-Hansen et al., 2007). For this reason, it is vital that multiple unlinked loci are investigated to form an accurate picture of the evolutionary history of populations of interest.

In this study, we examine population genetic patterns using sequence data from the dopamine receptor *D4* (*DRD4*) gene in introduced starling *Sturnus vulgaris* populations in Australia. Polymorphisms of *DRD4* have been associated with variation in personality traits in humans, but support for this relationship is equivocal (Kluger et al., 2002). Polymorphisms in exon 3 of this gene have also been linked to novelty-seeking behaviour in great tits *Parus major* (Fidler et al., 2007) and yellow-crowned bishops *Euplectes afer* (Mueller et al., 2014) but this gene has not been studied in starlings.

Starlings were introduced to eastern Australia (New South Wales, Victoria, Tasmania and South Australia) in the mid-19th Century (Jenkins, 1977) and have subsequently spread westward into Western Australia across the arid Nullarbor Plain (Fig. 1; Woolnough et al., 2005). Other than the occasional vagrant, starlings were first found in Western Australia in the area surrounding Condingup (Fig. 1) in 1971, presumably following the construction of a bitumen road across the arid expanse of the Nullarbor Plain, which may have increased access to water (Woolnough et al., 2005). In 2001, starlings were discovered approximately 200 km to the west of Condingup in the vicinity of Munglinup, Western Australia. It has been shown that these two incursions are likely to represent separate secondary introductions from different parts of eastern Australia (Rollins et al., 2009; Rollins et al., 2011).

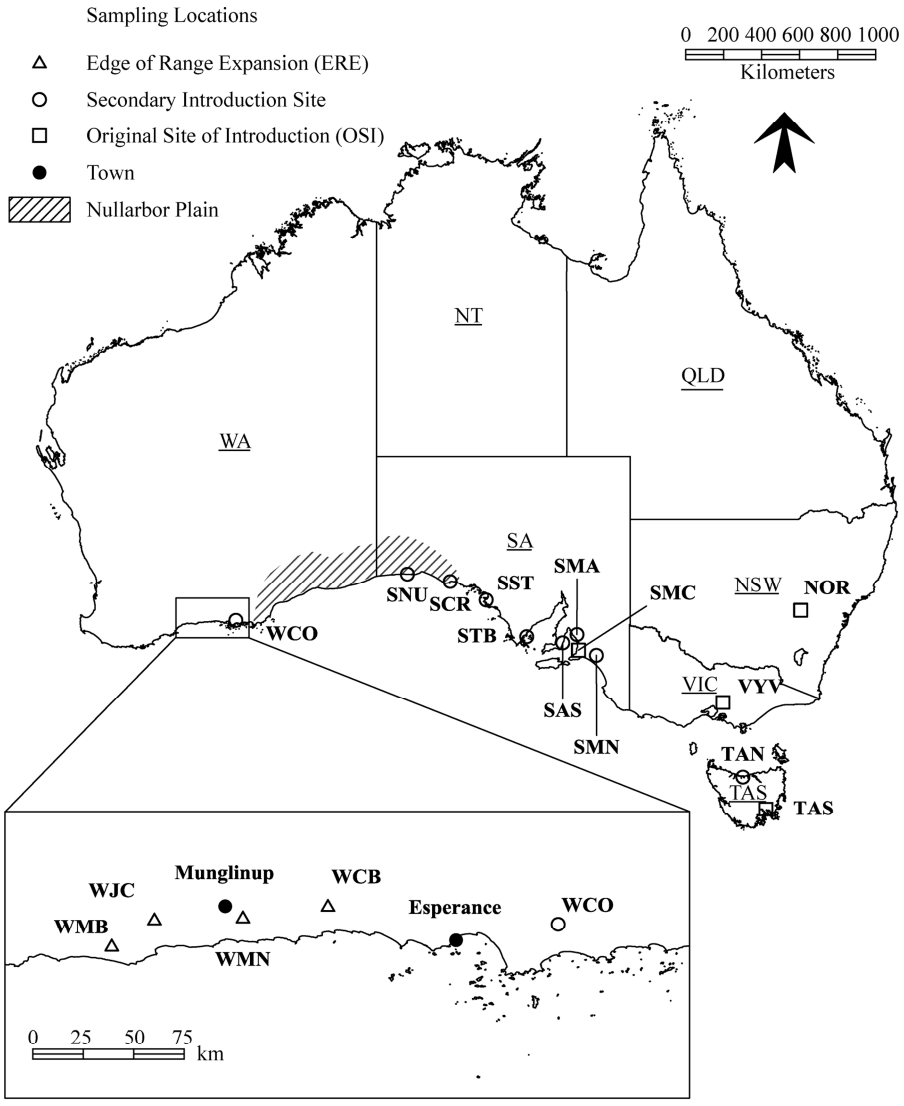
The well-documented history of invasion and availability of data from both nuclear (Rollins et al., 2009) and mitochondrial (Rollins et al., 2011) neutral markers for comparison, as well as the potential for strong selection in invasive populations, makes this an ideal study system to evaluate how well *DRD4* describes the evolu-

tionary history of invasive starlings in Australia. Here we combined *DRD4* gene sequence data with putatively neutral mtDNA and microsatellite data to address the following:

1) Range expansion populations in the current study may well have been established by individuals exhibiting novelty-seeking behaviour. We ask whether *DRD4* allele proportions differ between localities near the original introduction sites as compared to those on the range expansion front.

2) In the early stages of invasion, if long-distance dispersers experience selection on *DRD4,* gene diversity should be suppressed. However, as more individuals arrive, this may no longer occur. We ask whether non-neutral gene sequence data show differing levels of genetic diversity to that found in putatively neutral mtDNA and microsatellite loci.

3) In light of what we have learned about the nature of selection on *DRD4*, we ask how useful this locus is for revealing population differentiation and invasion history and if all of the markers assessed infer similar invasion histories.



WMB (N=7) Mason Bay, WA
WJC (N=21) Jerdacuttup, WA
WMN (N=23) Munglinup, WA
WCB (N=22) Coomalbidgup, WA
WCO (N=26) Condingup, WA
SNU (N=27) Nullarbor, SA
SCR (N=33) Coorabie, SA
SST (N=29) Streaky Bay, SA
STB (N=30) Tumby Bay, SA
SAS (N=29) Stansbury, SA
SMA (N=33) Mallala, SA
SMC (N=32) McLaren Vale, SA
SMN (N=30) Meningie, SA
VYV (N=29) Yarra Valley, VIC
NOR (N=30) Orange, NSW
TAN (N=17) Devonport, TAS
TAS (N=12) Hobart, TAS

Fig. 1 Map of sampling localities

Abbreviations and sample sizes (*N*) used for *DRD4* sequencing shown below map. Original sites of introduction (OSI) denoted with open squares. Inset map shows sites > 1,000 km from OSI.

1 Material and Methods

1.1 Sampling and DNA extraction

Samples were collected during control operations at 17 localities across the introduced range of starlings in Australia (Fig. 1). Some individuals were shot at nests (N = 230), some were trapped in modified crow traps containing lure birds (n = 113), and the method of capture of some birds was not recorded (n = 87). Samples from areas with low sample size were pooled with those from nearby locations (WMB with WJC, 32 km apart; TAS with TAN, 190 km apart; see Fig. 1), after *DRD4* data were tested for homogeneity in ARLEQUIN using p-values of pairwise exact test on allele frequencies. DNA was extracted using a Gentra PureGene extraction kit (QIAGEN) according to manufacturer's instructions. These same individuals were previously analysed for between-population variation using putatively neutral markers including eleven microsatellites (Rollins et al., 2009) and 942 base pairs (bp) of mitochondrial control region (mtCR) sequence (Rollins et al., 2011). We compare our current results to data from these previous studies (see below).

1.2 Primer design

Because exon sequence is typically conserved across taxa, primers to amplify introns are often designed in adjoining exons, termed EPIC (exon-primed intron-crossing) primers (Palumbi and Baker, 1994). We designed primers for this study using the EPIC approach and PRIMER3 (Rozen and Skaletsky, 2000) software. Primers D4E3a-F and D4E4a-R (Table 1) were designed in *DRD4* exons 3 and 4 respectively, using blocks of sequence conserved between the chicken genome and *P. major DRD4* (GenBank DQ006801). The resulting amplicon was approximately 700 bp in length and contained a portion of exon 3 and the entirety of the intron found between exons 3 and 4 (Fig. 2). Included in this sequence was the portion of exon 3 that Fidler et al. (2007) found to contain polymorphisms that were linked to exploratory behaviour in *P. major*. An NCBI BLAST search revealed this sequence matched well to *DRD4* sequences of a wide variety of passerine birds (e.g. *P. major*, DQ006801, query coverage = 100%, e-value = 0, sequence identity = 89%; *Taeniopygia guttata*, GQ359780.1, query coverage = 39%, e-value = 9e-93, sequence identify = 96%). We designed a second set of primers internal to the first set: D4E3b-F and D4I3a-R (Table 1, Fig. 2).

1.3 DRD4 amplification and sequencing

The initial polymerase chain reaction (PCR) for all samples was conducted using primer pair D4E3a-F / D4E4a-R. PCR reactions containing 30 ng of genomic DNA, 0.5 μM primers, 1x buffer, 2.5 mM MgCl, 200 μM dNTPs, and 0.2 units of AmpliTaq Red (ABI) in a final volume of 20 μl were run for 30 cycles at 59° C (annealing temperature). Samples with poor or no amplification were re-run using alternate primer combinations and 3.0 mM MgCl. All PCR products were cleaned with Exo-Sap-It (GE Healthcare) following manufacturer's instructions. Big-Dye chemistry (ABI) was used in sequencing reactions and products were ethanol precipitated and run on an ABI 3730. A subset of samples was sequenced in the reverse direction to assure accuracy of base identification. All individuals having singleton polymorphisms were sequenced twice to minimise sequencing error. Sequences were edited, aligned

Table 1 *Sturnus vulgaris DRD4* primer sequences and melting temperatures

Position	Primer	Sequence	T_m °C
Exon 3	D4E3a-F	5' – TGGAGGATGACAACTACATCG	61
Exon 3	D4E3b-F	5' – CATGTTCCAAGGACTCAAGC	60
Intron 3	D4I3a-R	5' – AAGGTGATGTTCAAAAGGGTT	57
Exon 4	D4E4a-R	5' – CCAGCAGAAGAGGAAAGCAC	62

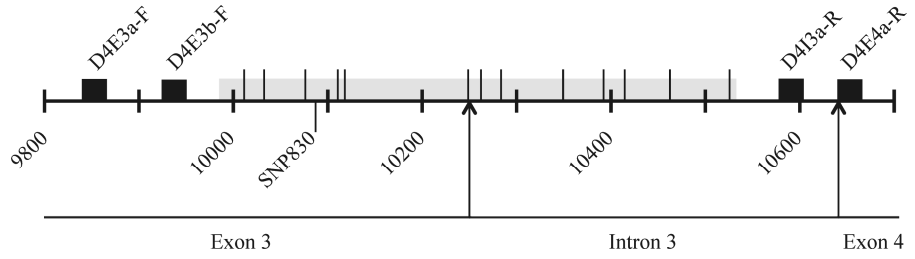


Fig. 2 Schematic of portion of *DRD4* gene sequenced in *S. vulgaris*

Primers from Table 1 are shown as black boxes. Sequence from Exon 3 and Intron 3 analysed here is highlighted in grey. SNPs are indicated by thin vertical lines. Information below the line relates to *DRD4* sequence from *P. major* (GenBank DQ006801), including SNP830, which showed a correlation with novelty-seeking behavior in that species. Exon/intron boundaries are shown at the bottom of the figure.

and trimmed to a uniform length in SEQUENCHER version 4.2. We aligned starling *DRD4* sequence with homologous sequence from the chicken and zebra finch genomes (phylogenetically closest published genomes) to delineate the boundary between exon 3 and intron 3. A data file was constructed containing both exon and intron data, as well as a file that only contained exon data for investigations of selection and one containing only intron data for comparisons of genetic diversity to other markers (see below).

1.4 Allele discrimination

We used the Excoffier-Laval-Balding (ELB) haplotype inference algorithm implemented in ARLEQUIN version 3.11 (Excoffier et al., 2005) to infer alleles (haplotypes). Excoffier et al. (2003) demonstrated that this method is much faster computationally than Bayesian methods, while performing equally well, and that in the presence of high levels of recombination, ELB's performance is more accurate. The ELB approach has been compared to other methods of haplotype inference (e.g. expectation-maximum and simple inference methods), all of which have yielded highly similar results (Zarowiecki et al., 2014). Values for α, ε, and γ were all set to 0.01 following recommendations of the authors (Excoffier et al., 2003). Because this method assumes Hardy-Weinberg equilibrium, data were tested for adherence to this assumption using ARLEQUIN. Using the combined dataset, allele proportions were calculated for each locality.

1.5 Tests for selection

Polymorphisms in the exon sequence were classified as synonymous or non-synonymous. For each polymorphism found, the number of copies of each genotype at that position were calculated for each of the following distance classes: localities at sites from the original introduction (NOR, VYV, TN/S, SMC), localities to the east of the Nullarbor Plain that were located less than 1000 km from introduction sites (SMN, SMA, SAS, STB, SST, SCR, SNU), and localities to the west of the Nullarbor Plain that were located more than 1,000 km from introduction sites (WCO, WCB, WMN, WM/J) (Fig. 1). Assuming long-distance dispersal is correlated with novelty-seeking behaviour in starlings as has been shown in great tits (Dingemanse et al., 2003), the rationale behind the choice of these groups was that if a sequence variant is associated with a predisposition for long-distance dispersal, the proportion of that variant should increase with distance from site of introduction. For each SNP, we used chi-square contingency tests to determine whether allele counts were different across distance classes. For each sampling locality, we used the combined dataset to calculate two neutrality indices: Tajima's *D* (1989) and Fu's *Fs* (1997). Although these indices are indicators of neutrality/selection, they are also sensitive to population expansion and contraction.

Patterns consistent with selection, identified using the methods described above, may also be the result of genetic drift. Therefore, it would be useful to have a measure of novelty-seeking behaviour that is independent to population history. Given the management of this invasive species, captive experiments were not possible. However, we did obtain samples using a variety of methods including shooting and trapping. Method of capture has been used previously as a proxy for boldness behaviour (González-Bernal et al., 2014) so we tested whether SNPs were associated with method of capture in our study. We used willingness to go into a trap as a proxy for boldness whereas individuals that were shot were used as controls.

1.6 Genetic diversity within populations

Values of *DRD4* intron 3 molecular diversity (θ_s) within populations were estimated in ARLEQUIN and plotted as a function of estimated time since colonization and a linear regression was performed on these data. To compare estimates of genetic diversity found in *DRD4* intron 3 to estimates from the same individuals using microsatellites (Rollins et al., 2009) and mtDNA data (Rollins et al., 2011), we selected a group of 318 individuals from 11 localities for which we had sufficient sample sizes for all three classes of markers (Table 2).

Table 2 Sample size and gene diversity (*GD*) calculated for the same group of individuals using microsatellite, *DRD4* and mtCR datasets

Locality	Sample Size	Microsatellite *GD*	*DRD4*-I3 *GD*	mtCR *GD*
WCO	25	0.630 (0.029)	0.868 (0.003)	0.680 (0.009)
SNU	27	0.639 (0.032)	0.915 (0.002)	0.704 (0.007)
SCR	31	0.666 (0.028)	0.915 (0.002)	0.740 (0.005)
SST	28	0.673 (0.031)	0.902 (0.003)	0.664 (0.007)
STB	28	0.655 (0.029)	0.892 (0.003)	0.704 (0.008)
SAS	26	0.648 (0.030)	0.896 (0.003)	0.668 (0.010)
SMA	32	0.641 (0.028)	0.914 (0.002)	0.692 (0.007)
SMC	32	0.664 (0.030)	0.889 (0.003)	0.692 (0.007)
SMN	30	0.675 (0.027)	0.857 (0.003)	0.639 (0.009)
VYV	29	0.655 (0.022)	0.820 (0.006)	0.569 (0.009)
NOR	30	0.650 (0.028)	0.941 (0.001)	0.800 (0.006)

Values in parentheses are standard error

Using this group of individuals, gene diversity (*GD*) was calculated in ARLEQUIN using *DRD4* data. For an equitable comparison, genetic diversity was calculated using microsatellites (equivalent to expected heterozygosity, H_E) and mtCR sequence data from these same individuals following methods in the previous studies of these markers (Rollins et al., 2009; Rollins et al., 2011). Linear regressions were performed in SPSS comparing *GD* of *DRD4* to that of microsatellite and mtCR data. Absolute amounts of *GD* were plotted for the three marker classes across all localities and paired *t*-tests were used to assess the significance of any differences found between *DRD4 GD* and similar estimates from the other two marker classes.

1.7 Genetic variation between populations

Genetic variation between populations was assessed using the combined dataset (exon and intron sequence). *P*-values from pairwise exact tests (Raymond and Rousset, 1995) were used to identify variation between localities and multiple comparisons were adjusted for false discovery rate (FDR) following the method of Benjamini and Hochberg (1995). AMOVA was used to examine hierarchical population structure based on the regions which were identified using microsatellite loci in this species (Rollins et al., 2009).

We also assessed population subdivision using the Bayesian clustering approach implemented in STRUCTURE version 2.3.4 (Pritchard et al., 2000; Falush et al., 2003). To analyse STRUCTURE results, we used the ΔK method (Evanno et al., 2005) to infer the number of genetic groups sampled. We chose the admixture model with correlated allele frequencies and ran ten replicates of each run and set the number of genetic groups to test to $K = 1$ to 8. Each run consisted of a burn-in period of 100,000 MCMC (Markov chain Monte Carlo) steps followed by 10^6 iterations.

2 Results

P-values of pairwise exact tests of *DRD4* differentiation between WMB and WJC ($P = 0.115$), and TAN and TAS ($P = 0.110$) were not significant. In all subsequent analyses, WMB and WJC are pooled (hereafter referred to as WM/J), as are TAN and TAS (TN/S).

2.1 Amplification and allele discrimination

The majority of samples used in this study amplified well with primer pair D4E3b-F / D4I3a-R. However, approximately one-quarter of our samples did not amplify well with this primer pair, presumably due to polymorphisms in the primer annealing sites. These samples did amplify using alternative primer pairs (Table 1).

The combined dataset for subsequent analyses contained 248 bp from the 3' end of exon 3, and 307 bp from the 5' end of intron 3. The exon sequence contained six single nucleotide polymorphisms and the intron sequence contained seven. None of these loci showed significant departures from Hardy-Weinberg equilibrium. Using the ELB algorithm, we defined thirty-eight alleles (Table 3; Supplementary Material, Table S1; GenBank KP941028).

2.2 Tests for selection

All six polymorphisms found in the exon sequence were third-codon position synonymous substitutions. Allele counts for four single-nucleotide polymorphisms (SNPs; sites 26, 56, 92 and 131) were significantly different across distance classes (Table 4). However, Tajima's *D* and Fu's *Fs* were not significantly different from zero for any sampling locality after FDR correction (Tajima's *D* ranged from 0.98–1.97; Fu's *Fs* ranged from -5.65 to -0.01). Further, none of the six SNPs identified in *DRD4* exon 3 were significantly associated with capture method (Table 4).

2.3 Genetic diversity within populations

DRD4 showed a significant linear relationship between date of first record and the nucleotide diversity measure θ_s ($r^2 = 0.55$, $P = 0.002$; Fig. 4). Levels of genetic diversity in localities from the westernmost localities (WM/J, WMN, WCB and WCO) were lower than those in other parts of Australia (Fig. 5). As a group, these western localities had lower values of θ_s (mean $\theta_s = 1.85$) than all other localities to the east (mean $\theta_s = 2.40$) (*t* test, $P < 0.001$). The locality having the greatest number of alleles was SMA (20 alleles) and the locality having the fewest was WCB (9 alleles). Using the subset of samples for which we had sufficient sample sizes to compare estimates across markers (see above), we calculated *GD* of microsatellites, *DRD4* and mtCR (Table 2). The regression of *DRD4 GD* versus microsatellite *GD* was not significant ($r^2 = 0.021$, $P = 0.670$). However, the regression of *DRD4 GD* versus mtCR *GD* was highly significant ($r^2 = 0.809$, $P < 0.001$). Across all sampling localities included in our subset of samples, estimates of *GD* were significantly lower for mtCR (*GD* ranged from 0.57 to 0.80; paired *t*-test $P < 0.001$) and microsatellite data (*GD* ranged from 0.63 to 0.67; paired *t*-test $P < 0.001$) than for *DRD4* intron data (*GD* ranged from 0.82 to 0.94; Fig. 6).

2.4 Genetic variation between populations

Large shifts in proportions of *DRD4* alleles 3, 5, 6 and 10 occurred across the range of starlings in Australia and all of these frequency changes occurred between

Table 3 ***DRD4*** **allele proportions for each sampling locality.**

Allele	WM/J*	WMN*	WCB*	WCO*	SNU	SCR	SST	STB	SAS	SMA	SMC	SMN	VYV	NOR	TN/S
1	0.054	0.043	0.068	**0.173**	0.074	0.091	0.052	0.050	0.034	0.091	**0.109**	0.083	**0.103**	0.050	0.052
2	**0.321**	**0.348**	**0.455**	0.058	**0.167**	**0.121**	0.069	**0.283**	**0.207**	0.076	**0.109**	**0.133**		0.083	**0.155**
3	0.036				**0.148**	**0.212**	**0.241**	**0.100**	0.086	**0.227**	**0.234**	**0.200**	0.017	0.067	**0.138**
4	0.018				0.019	0.015				0.015					
5	**0.179**	**0.174**	**0.114**	**0.115**	0.056	0.045	0.017	0.033	0.017			0.017	0.034		
6	**0.196**	**0.152**	0.091	**0.212**	0.071	0.030	0.052	0.017	0.017	0.045		0.017	0.052	0.033	0.034
7	0.089	0.065	0.045	0.038		0.015				0.015			0.017		0.017
8	0.054	0.109	0.159	0.019	0.037	0.030	0.017	0.017	0.069	0.030	0.031		0.069	0.067	
9	0.018	0.043				0.076	0.069	0.033	**0.121**	0.061	0.047	**0.100**	0.086	0.067	0.069
10		0.022		**0.173**	**0.148**	**0.121**	**0.190**	**0.117**	**0.172**	**0.121**	**0.172**	**0.267**	**0.397**	**0.133**	**0.293**
11		0.022													
12		0.022	0.023												
13			0.023	**0.154**	0.054	0.030	0.034	0.033		0.015			0.017	0.050	0.017
14			0.023	0.019			0.034			0.015	0.016			0.050	
15				0.019											
16	0.018			0.019	0.036	0.030	0.069	0.067	0.017	0.076	0.047	0.033	0.034	0.017	0.086
17					0.018	0.030				0.015	0.047			0.033	
18	0.018				0.071	0.030	0.052	**0.117**	0.069		0.016	0.017	0.017	0.083	0.034
19					0.054	0.030	0.017	0.017	**0.103**		0.047		0.017		0.052
20					0.018										
21					0.018										
22						0.045		0.050	0.017	0.061	0.063	0.050		**0.117**	
23						0.030	0.017		0.052	0.045	0.047	0.083		0.033	0.017
24						0.015				0.015					
25							0.034						0.069		
26							0.017			0.015					
27							0.017			0.015				0.017	
28								0.017							0.017
29								0.017							
30								0.017							
31								0.017							
32									0.017						
33										0.015					
34										0.030	0.016		0.052	0.067	
35													0.017		
36														0.017	
37														0.017	
38															0.017

Common alleles for each population (frequency > 0.100) in bold. SNPs for each allele are shown in Supplementary Material, Table S1. Full sequence deposited in GenBank KP941028. Localities are organised from west to east and those > 1000 km from the original sites of introduction are denoted by asterisks

Table 4 Allele counts per substitution site in *DRD4* exon data

	Site 26		Site 56		Site 92		Site 125		Site 131		Site 248	
	G	*C*	*G*	*C*	*T*	*C*	*G*	*A*	*T*	*C*	*T*	*C*
Distance class												
1 (OSI)	66	174	235	5	136	104	233	7	203	37	240	0
2 (< 1,000 km from OSI)	131	291	420	2	285	137	409	13	360	62	418	4
3 (> 1,000 km from OSI)	123	75	198	0	274	24	197	1	136	62	198	0
χ^2	69.15		7.02		50.64		4.07		26.86		4.17	
df	2		2		2		2		2		2	
P	< 0.001		0.030		< 0.001		0.131		< 0.001		0.124	
Capture method												
Shot	179	281	453	7	314	146	452	8	371	89	457	3
Trapped	85	141	226	0	170	56	220	6	182	44	225	1
χ^2	0.06		2.13		3.21		0.26		0.00		0.04	
df	1		1		1		1		1		1	
P	0.807		0.144		0.072		0.610		1.000		0.842	

Distance class: data are separated into three distance classes from original sites of introduction (OSI). Capture method: data are separated according to whether the sample was shot or trapped. Chi square values, degrees of freedom (*df*) and *P*-values are given.

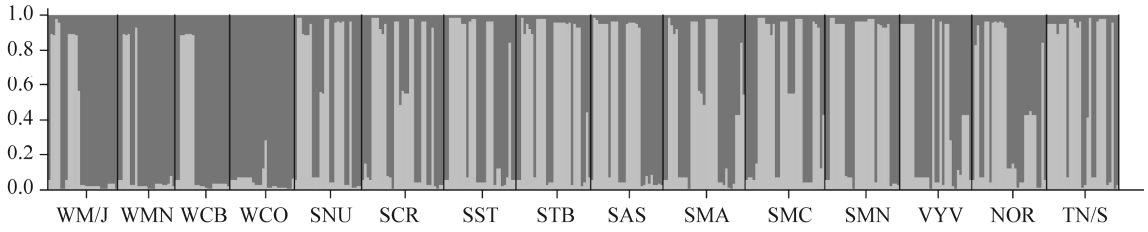


Fig. 3 STRUCTURE Q plot showing membership of two genetic groups present in Australia, based on sequence data from *DRD4*

Each individual is represented by a vertical bar indicating degree of admixture.

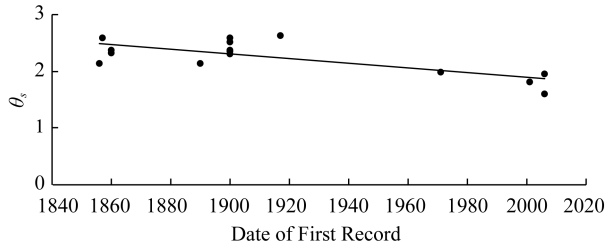


Fig. 4 Relationship between molecular diversity (θ_s) and date of first starling record (r^2 = 0.550, P = 0.002)

WCO and one of its two adjoining localities (Table 3). Exact tests indicated that the three westernmost sampling localities (WM/J, WMN, WCB) were not differentiated from one another (*P*-values ranged from 0.187 to 0.851; Supplementary Material, Table S2) but were significantly differentiated from all other sampling localities across Australia (all *P*-values < 0.001). After FDR correction for multiple comparisons, two sites were significantly differentiated from all other localities sampled (WCO and VYV; all pairwise *p*-values with these sites were < 0.001; Supplementary Material, Table S2). Using the four regions defined by microsatellite data (eastern Australia including NSW, TAS, VIC; all sites in South Australia; WCO in Western Australia; WMB, WJC, WMN, WCB in Western Australia; Rollins et al., 2009), AMOVA of *DRD4* data showed significant differentiation among regions (8.3% of variation, *P*-value < 0.001, F_{CT} = 0.083) but the majority of the variation was found among localities among regions (91.6%, *P*-value < 0.001, F_{ST} = 0.084). Less than 1% of the variation was attributed to differences among localities within regions (*P*-value = 0.309, F_{SC} = 0.002). STRUC-

TURE analyses indicated the presence of two genetic groups: all Western Australian localities formed one group and all other localities formed the other group (Fig. 3).

3 Discussion

3.1 Evidence of selection

Polymorphisms of the DRD4 gene have been linked to novelty-seeking behaviour in humans (Chen et al., 1999), where it has been shown to be highly variable (Ding et al., 2002). *DRD4* may provide one of the best-described examples of balancing selection in humans (Ding et al., 2002). However, investigations of the relationship between *DRD4* alleles and such behaviours have not always yielded clear results (Kluger et al., 2002). More recently, this gene has been investigated in other taxa in relation to behaviour. Here we ask whether an invasive population of starlings displays evidence of selection at this locus. Contrary to the results of *DRD4* studies in great tits (Fidler et al., 2007) and yellow-crowned bishops (Mueller et al., 2014), our investigations did not find compelling evidence of selection for novelty-seeking behaviour acting on this locus in Australian starling populations. No significant departures from Hardy-Weinberg equilibrium were found. Neither Tajima's *D* nor Fu's *Fs* showed evidence of selection for any locality but these tests will only detect selection at specific times relative to its onset and the time at which equilibrium is regained, which will vary depending upon the type and intensity of selection (Halliburton, 2004). All identified substitutions were synonymous (i.e. did not result in an amino acid change). However, evidence now suggests that synonymous mutations can alter translation and therefore gene expression, and this has been demonstrated in the *DRD2* gene in humans (Duan et al., 2003).

We did find three polymorphisms showing differences in relative abundance across distance classes (Table 4),

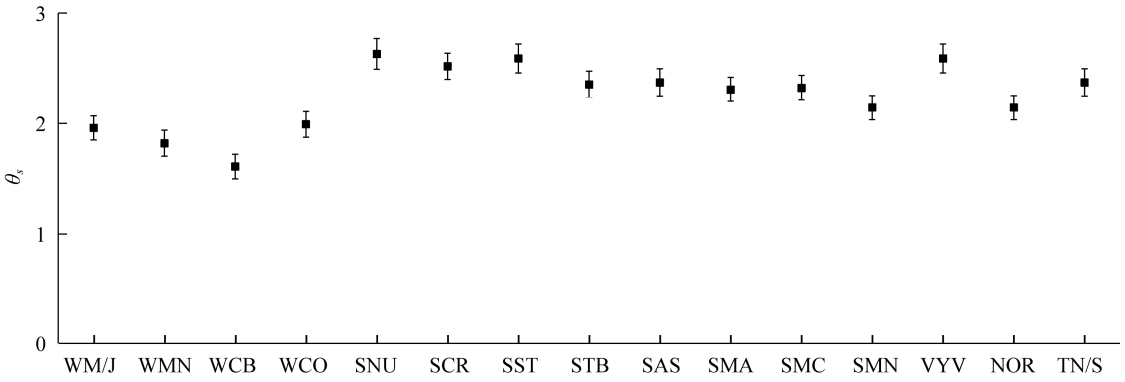


Fig. 5 Molecular diversity (θ_s) found in each sampling locality
Error bars represent standard error. Localities are organised from west to east.

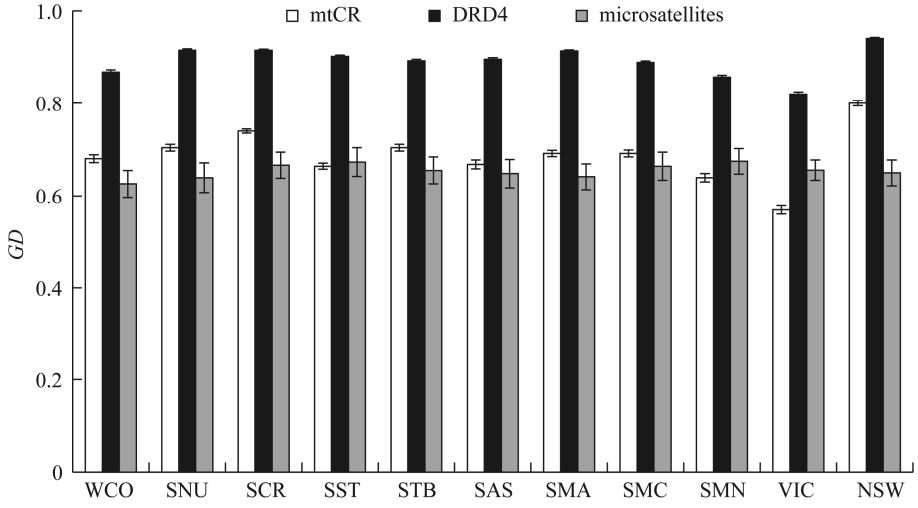


Fig. 6 Gene diversity (*GD*) estimated from mtCR, DRD4 (intron 3) and microsatellite data with standard error bars
Localities are organised from west to east.

which might be linked to novelty-seeking behaviour. However, these changes could also be the result of genetic drift. To investigate association between these SNPs and novelty-seeking behaviour, we analysed these data according to whether individuals were shot (control sample) or trapped using conspecific lure birds. None of these SNPs were significantly associated with capture method, suggesting that drift may be the more likely explanation for allele differences across populations. In summary, we have found little support for selection acting on *DRD4* in Australian starling populations.

Given results of other studies of *DRD4* in birds that have found evidence of selection at this locus (Fidler et al., 2007; Mueller et al., 2014), our results may be surprising. However, it has previously been demonstrated that Australian starling populations show evidence of genetic bottlenecks (Rollins et al., 2009) and substantially reduced genetic diversity as compared to starlings in their native range (Rollins et al., 2011). Therefore, diversity of functional polymorphisms at *DRD4* could have been lost through the process of introduction. Further, it is possible that capture methods used when starlings were translocated from their native range resulted in pre-establishment selection (Carrete et al., 2012) for particular personality types, and potentially for particular *DRD4* alleles. It is also possible that the earliest immigrants to Western Australia were bolder than later immigrants and these early immigrants are no longer represented in the population we sampled. Finally, we have only sequenced a portion of this gene, and other coding regions of *DRD4* may harbour functional polymorphisms in starlings. Further support for an absence of directional selection at *DRD4* comes from our investigations of within-population genetic variation (below).

It should be noted that Korsten et al. (2010) conducted a second study to investigate the relationship between novelty-seeking behaviour and *DRD4* in free-living great tits from four wild populations, including that used for Fidler et al.'s (2007) study (Westerheide). While the same patterns emerged from Westerheide samples in both studies, the remaining three populations in Korsten et al. (2010)'s study showed little evidence of selection for exploratory behaviour at this locus. They point out that at Westerheide, only approximately 5% of the variation in exploratory behaviour was explained by genetic variation at *DRD4* and, in a meta-analysis using human data, genetic variation at this locus only explains about 3% of novelty-seeking behaviour (Munafò et al., 2008). These small effects are, nonetheless, substantial by the standards of other association studies of complex traits (Munafò and Flint, 2004). However, such small effects may be difficult to detect. Further in-depth investigation of *DRD4* in these same four populations revealed evidence of association between exploratory behaviour and different SNPs across populations within this gene, indicating that selection for exploratory behaviour may have evolved locally in these populations (Mueller et al., 2013). Therefore, it is possible that in Australian starlings, other areas of this gene outside of our sequenced region may contain SNPs linked to exploratory behaviour.

3.2 Genetic diversity within populations

Other than singleton alleles, no *DRD4* allele was private to any sampling locality, nor to any region pre viously defined using microsatellite data (Rollins et al., 2009). To the east of the Nullarbor Plain, molecular diversity (θ_s) was similar across all sampling localities, and at localities in WA, θ_s was lower. Because θ_s is related to effective population size, we can infer that localities in WA might have experienced smaller effective population sizes than those in the east; this pattern could also result from directional selection occurring in WA localities, but in that case we would expect to see lowered diversity only at *DRD4* (and not putatively neutral loci), which is not the pattern we observed. The lowest value of molecular diversity of *DRD4* was found in WCB, in accordance with previous results from both the microsatellite and mitochondrial datasets (Rollins et al., 2009; Rollins et al., 2011). Similarly, molecular diversity measures for all three marker types were lower for sampling localities with a more recent date of first record (e.g. Western Australian localities, Figs. 4–6), supporting the idea that this measure may be useful for determining whether newly founded populations result from recent incursions.

Väli et al. (2008) investigated the relationship between diversity found at microsatellite loci and diversity found at random DNA sequences, studying four species of carnivores from a total of eight populations. They found that nucleotide diversity estimated from a suite of 10 intron sequences varied 30-fold across their study populations. Their estimate of microsatellite gene diversities for the same individuals varied only 1.4-fold (or 3-fold when converted to $H_e/(1-H_e)$, which is directly comparable to nucleotide diversity; Väli et al., 2008). The large difference in diversity found across the two marker classes used by Väli et al. (2008) was driven by the large number of introns having no variation in two populations included in their study, which are believed

to have passed through a bottleneck (Hellborg et al., 2002; Cegelski et al., 2006). However, microsatellite gene diversity in their populations (approximately 0.50) did not reflect the extreme lack of intron diversity. This could be because of the different ways that variable microsatellites were ascertained in the studies summarised by Väli et al. (2008), or possibly because the microsatellite diversity had recovered from the bottleneck faster, due to a higher mutation rate (Halliburton, 2004). In our study, within-population diversity estimates (*GD*) using the *DRD4* data had a larger range across localities (1.15-fold) than *GD* estimates using microsatellite data (1.07-fold). The difference was much less pronounced than in the Väli et al., (2008) study, possibly because they included several species, while our study included only one species, using populations resulting from a common introduction.

While it is expected that diversity measures across different marker classes should be correlated in the absence of selection, in our study no significant relationship was found between *DRD4* and microsatellite gene diversity. This may be due to ascertainment bias because microsatellites are typically chosen by virtue of their degree of diversity. As a result, while populations may have varying degrees of diversity at *DRD4*, they may have similar diversity estimates using microsatellite loci, which were originally chosen because they showed consistently high levels of diversity across populations. However, we found a highly significant relationship between *DRD4* and mitochondrial gene diversity ($r^2 = 0.808$, $P < 0.001$). In a study of medfly native and invasive populations, Villablanca et al. (1998) found a greater loss of gene diversity at introduction in mitochondrial data than at nuclear intron data. In their study, for both native and introduced populations, nuclear gene diversity was higher than mitochondrial gene diversity. Similarly, in our study, gene diversity was significantly greater at *DRD4* than for mtCR or microsatellite data, in every locality tested. This pattern could result from higher mutation rates or balancing selection at *DRD4*, increasing diversity at this locus, or from a reduction in diversity at other loci. For example, Villablanca et al. (1998) concluded that the greater loss of diversity in the mitochondrial genome at introduction was likely to be a result of stronger genetic drift acting on the mitochondrial genome, whose inheritance mode lowers the effective population size relative to nuclear DNA.

We note that our investigations of within-population genetic variation also support an absence of directional selection at *DRD4*. Although newly-established invasive populations are not likely to be at equilibrium, Chao et al. (2015, in press) have demonstrated that mutation-drift equilibrium is a reasonable approximation for Australian starlings. Assuming equilibrium, within-population *GD* is proportional to $4N_e\mu$ / $(1+4N_e\mu)$, where N_e is the effective population size and μ is the mutation rate (Crow and Kimura 1970) for loci evolving under the Infinite Alleles Model (e.g. *DRD4*, *mtCR*). This relationship is similar for loci evolving under the Stepwise Mutation Model (e.g. microsatellites, Kimura and Ota, 1975). Mutation rates have been estimated for nuclear genes at 10^{-5} per gene of ~1,000 bases, mitochondrial DNA an order of magnitude higher, and microsatellites at 10^{-3} (Halliburton, 2004); all these rates might vary by an order of magnitude. Effective population sizes in starlings might be between 1,000 to 10,000 (Rollins unpublished). Effective population size of mtDNA is four-fold lower than nuclear markers (Ballard and Whitlock, 2004). Applying these values in the equation above show that unless population sizes are extremely small, changes to N_e have a small effect on estimates of *GD*, whereas differences in estimates of μ across markers impacts *GD*. If we assume all of these loci are neutral, we would expect that *GD DRD4* should be somewhat lower than for mtDNA and microsatellites. However, our data show that *GD* estimated from *DRD4* is clearly not lower than *GD* estimated from mtDNA or microsatellite loci, but actually slightly higher (Fig. 6). Thus there is not perfect agreement to a neutral scenario, but certainly not evidence for directional selection, which would be expected to depress within-population variability of *DRD4* relative to neutral expectations (Nielsen, 2005).

3.3 Genetic variation between populations

Using the *DRD4* dataset, we found similar patterns of population differentiation to that found for the same individuals using microsatellites (Rollins et al., 2009) and mtDNA sequence data (Rollins et al., 2011). Localities in WA were differentiated from eastern localities (Fig. 3), and exact tests indicated that WA localities on the edge of the range expansion (WM/J, WMN, WCB) were not different to one another but were differentiated from the original area of colonisation in WA (WCO). Similar to what was found with microsatellite and mtCR data in previous studies of these individuals, shifts in allele proportions occurred on either side of WCO (Table 3), indicating that this may be an area where two genetic groups are converging. Although it appears that broad patterns are similar across the markers we used in this study, we found a greater degree of between-population

variation using *DRD4* (Supplementary Material, Table S2), possibly due to the higher number of alleles found at this locus. Within the two eastern regions defined by microsatellites (Rollins et al., 2011), exact tests indicated that some localities were differentiated from others at *DRD4*. In particular, VYV was differentiated from all other localities in the East. Allele 10 was the most common allele only in the four easternmost localities (TN/S, NOR, VYV, SMN; Table 3) and the frequency of this allele was extremely high in VYV (0.397). It is possible that this reflects the genetic composition of the founding population in VYV. Although substructure was identified within the regions delineated by microsatellite data, in the *DRD4* dataset more of the variation could be explained by differences between regions (8.36%) than with the microsatellite dataset (4.6%). This suggests that the *DRD4* dataset may have a greater power to identify variation between populations.

3.4 Nuclear gene sequences and invasive population genetics

One of the greatest challenges in using genetics to understand the history and dynamics of invasive populations stems from the reduction in genetic diversity that often accompanies newly established populations. Villablanca et al. (1998) concluded that markers with high levels of genetic diversity are likely to be more useful to the study of invasive populations (Villablanca et al., 1998). In the present study, *DRD4* showed a greater degree of population differentiation than either microsatellites or mtDNA sequence. While we do not suggest that sequence data from a single nuclear gene is representative of an entire genome, the results from this study and others (Villablanca et al., 1998; Pampoulie et al., 2006; Hemmer-Hansen et al., 2007; Väli et al., 2008) indicate that the use of functional loci may provide more resolution than neutral markers traditionally used for population studies.

It is important to note that using loci under strong selection in population genetic studies requires careful interpretation because most analyses assume a neutral model of evolution. On the other hand, choosing loci believed to be under weak or no selection may avoid these problems, but these loci may also be uninformative. Recombination, mutation, effective population size and selection might impact each gene differently, and the history of a particular gene may not reflect the history of the genome. As a result, it is vital to compare multiple loci to determine whether results are indicative of population history.

If comparative reductions in within-population genetic diversity are used to determine whether newly discovered invasions are newly founded, it will be important to choose markers which maximise differences in diversity across populations. Estimates of diversity in this study were more variable across localities using *DRD4* data than with microsatellites, and Väli et al. (2008) found a 30-fold difference in diversity when comparing nuclear sequence data to microsatellites. We suggest that sequence data from functional nuclear genes can provide a valuable contribution to the study of invasive populations, especially if used in conjunction with other more traditional approaches, for which analyses are better developed and understood.

Acknowledgements We thank field staff at DAFWA and Nick Mooney for assistance with sampling, Scott Edwards for his suggestion to investigate *DRD4* in the context of invasion, and the Ramaciotti Centre for Gene Function Analysis for assistance with Sanger sequencing. Funding for this project was provided by Australian Research Council Linkage Grant #LP0455776. LAR is currently supported by Deakin University fellowship.

References

Akey JM, Zhang G, Zhang K, Jin L, Shriver MD, 2002. Interrogating a high-density SNP map for signatures of natural selection. Genome Research 12: 1805–1814.

Allendorf FW, England PR, Luikart G, Ritchie PA, Ryman N, 2008. Genetic effects of harvest on wild animal populations. Trends in Ecology & Evolution 23: 327–337.

Andolfatto P, 2005. Adaptive evolution of non-coding DNA in *Drosophila*. Nature 437: 1149–1152.

Ballard JWO, Whitlock MC, 2004. The incomplete natural history of mitochondria. Molecular Ecology 13: 729–744.

Benjamini Y, Hochberg Y, 1995. Controlling the false discovery rate: A practical and powerful approach to multiple testing. Journal of the Royal Statistical Society. Series B, Methodological. 57: 289–300.

Bowcock AM, Kidd JR, Mountain JL, Hebert M, Carotenuto L et al., 1991. Drift, admixture, and selection in human evolution: A study with DNA polymorphisms. Proceedings of the National Academy of Sciences of the United States of America 88: 839–843.

Carrete M, Edelaar P, Blas J, Serrano D, Potti J et al., 2012. Don't neglect pre-establishment individual selection in deliberate introductions. Trends in Ecology & Evolution 27: 67–68.

Cavalli-Sforza LL, 1966. Population structure and human evolution. Proceedings of the Royal Society of London. Series B: Biological Sciences 164: 362–379.

Cegelski CC, Waits LP, Anderson NJ, Flagstad O, Strobeck C et al., 2006. Genetic diversity and population structure of wolverine *Gulo gulo* populations at the southern edge of their current distribution in North America with implications for genetic viability. Conservation Genetics 7: 197–211.

Chao A, Jost L, Hsieh TC et al., 2015, in press. Expected Shannon entropy and Shannon differentiation between subpopulations

for neutral genes under the finite island model. PLoS ONE.

Chen C, Burton M, Greenberger E, Dmitrieva J, 1999. Population migration and the variation of dopamine D4 receptor (DRD4) allele frequencies around the globe. Evolution and Human Behavior 20: 309–324.

Crow JF, Kimura M, 1970. An Introduction to Population Genetics Theory. New York, Evanston and London: Harper & Row.

Ding Y-C, Chi H-C, Grady DL, Morishima A, Kidd JR et al., 2002. Evidence of positive selection acting at the human dopamine receptor D4 gene locus. Proceedings of the National Academy of Sciences 99: 309–314.

Dingemanse NJ, Both C, van Noordwijk AJ, Rutten AL, Drent PJ, 2003. Natal dispersal and personalities in great tits *Parus major*. Proceedings of the Royal Society of London. Series B: Biological Sciences 270: 741–747.

Duan J, Wainwright MS, Comeron JM, Saitou N, Sanders AR et al., 2003. Synonymous mutations in the human dopamine receptor D2 (DRD2) ffect mRNA stability and synthesis of the receptor. Human Molecular Genetics 12: 205–216.

Evanno G, Regnaut S, Goudet J, 2005. Detecting the number of clusters of individuals using the software structure: A simulation study. Molecular Ecology 14: 2611–2620.

Excoffier L, Laval G, Balding D, 2003. Gametic phase estimation over large genomic regions using an adaptive window approach. Human Genomics 1: 7–19.

Excoffier L, Laval G, Schneider S, 2005. Arlequin ver. 3.0: An integrated software package for population genetics data analysis. Evolutionary Bioinformatics Online 1: 47–50.

Falush D, Stephens M, Pritchard JK, 2003. Inference of population structure using multilocus genotype data: Linked loci and correlated allele frequencies. Genetics 164: 1567–1587.

Fidler AE, van Oers K, Drent PJ, Kuhn S, Mueller JC et al., 2007. Drd4 gene polymorphisms are associated with personality variation in a passerine bird. Proceedings of the Royal Society of London. Series B: Biological Sciences 274: 1685–1691.

González-Bernal E, Brown GP, Shine R, 2014. Invasive cane toads: Social facilitation depends upon an individual's personality. PLoS ONE 9: e102880.

Halliburton R, 2004. Introduction to Population Genetics. Upper Saddle River, N.J.: Pearson Education, Inc.

Hellborg L, Walker CW, Rueness EK, Stacy JE, Kojola I et al., 2002. Differentiation and levels of genetic variation in northern European lynx *Lynx lynx* populations revealed by microsatellites and mitochondrial DNA analysis. Conservation Genetics 3: 97–111.

Hemmer-Hansen J, Nielsen EE, Frydenberg J, Loeschcke V, 2007. Adaptive divergence in a high gene flow environment: Hsc70 variation in the European flounder *Platichthys flesus* L.. Heredity 99: 592–600.

Jenkins CFH, 1977. The Noah's Ark Syndrome. The Zoological Gardens Board, Western Australia.

Karl SA, Avise JC, 1992. Balancing selection at allozyme loci in oysters: Implications from nuclear RFLPs. Science 256: 100–102.

Kimura M, Ota T, 1975. Distribution of allelic frequencies in a finite population under stepwise production of neutral alleles. Proceedings of the National Academy of Sciences 72: 2761–2764.

Kluger AN, Siegfried Z, Ebstein RP, 2002. A meta-analysis of the association between *DRD4* polymorphism and novelty seeking. Molecular Psychiatry 7: 712–717.

Korsten P, Mueller JC, Hermannstädter C, Bouwman KM, Dingemanse NJ et al., 2010. Association between DRD4 gene polymorphism and personality variation in great tits: A test across four wild populations. Molecular Ecology 19: 832–843.

Lewontin RC, Krakauer J, 1973. Distribution of gene frequency as a test of the theory of the selective neutrality of polymorphisms. Genetics 74, 175–195.

Mueller J, Edelaar P, Carrete M, Serrano D, Potti J et al., 2014. Behaviour-related DRD4 polymorphisms in invasive bird populations. Molecular Ecology 23: 2876–2885.

Mueller JC, Korsten P, Hermannstaedter C, Feulner T, Dingemanse NJ et al., 2013. Haplotype structure, adaptive history and associations with exploratory behaviour of the *DRD4* gene region in four great tit *Parus major* populations. Molecular Ecology 22: 2797–2809.

Munafò MR, Flint J, 2004. Meta-analysis of genetic association studies. Trends in Genetics 20: 439–444.

Munafò MR, Yalcin B, Willis-Owen SA, Flint J, 2008. Association of the dopamine D4 Receptor (*DRD4*) gene and approach-related personality traits: Meta-analysis and new data. Biological Psychiatry 63: 197–206.

Nielsen R, 2005. Molecular signatures of natural selection. Annual Review of Genetics 39: 197–218.

Palumbi SR, Baker CS, 1994. Contrasting population structure from nuclear intron sequences and mtDNA of humpback whales. Molecular Biology and Evolution 11: 426–435.

Pampoulie C, Ruzzante DE, Chosson V, Jörundsdóttir TD, Taylor L et al., 2006. The genetic structure of Atlantic cod *Gadus morhua* around Iceland: Insight from microsatellites, the *Pan* I locus, and tagging experiments. Canadian Journal of Fisheries and Aquatic Sciences 63: 2660–2674.

Pimpinelli S, Dimitri P, 1989. Cytogenetic analysis of segregation distortion in *Drosophila melanogaster*: The cytological organization of the Responder (Rsp) locus. Genetics 121: 765–772.

Pritchard JK, Stephens M, Donnelly P, 2000. Inference of population structure using multilocus genotype data. Genetics 155: 945–959.

Raymond M, Rousset F, 1995. An exact test for population differentiation. Evolution 49: 1280–1283.

Rollins L, Woolnough AP, Sinclair RG, Mooney NJ, Sherwin WB, 2011. Mitochondrial DNA offers unique insights into invasion history of the common starling. Molecular Ecology 20: 2307–2317.

Rollins LA, Woolnough AP, Wilton AN, Sinclair RG, Sherwin WB, 2009. Invasive species can't cover their tracks: Using microsatellites to assist management of starling *Sturnus vulgaris* populations in Western Australia. Molecular Ecology 18: 1560–1573.

Rozen S, Skaletsky H, 2000. Primer3 on the WWW for general users and for biologist programmers. In: Krawetz S, Misener S ed. Bioinformatics Methods and Protocols: Methods in Molecular Biology. Totowa, NJ: Humana Press, 365–386.

Väli Ü, Einarsson A, Waits L, Ellegren H, 2008. To what extent do microsatellite markers reflect genome-wide genetic diversity in natural populations? Molecular Ecology 17: 3808–3817.

Villablanca FX, Roderick GK, Palumbi SR, 1998. Invasion

genetics of the Mediterranean fruit fly: Variation in multiple nuclear introns. Molecular Ecology 7: 547–560.

Woolnough AP, Massam MC, Payne RL, Pickles GS, 2005. Out on the border: Keeping starlings out of Western Australia. 13th Australasian Vertebrate Pest Conference, Wellington, NZ, 183–189.

Zarowiecki M, Linton YM, Post RJ, Bangs MJ, Htun PT et al., 2014. Repeated landmass reformation limits diversification in the widespread littoral zone mosquito *Anopheles sundaicus sensu lato* in the Indo-Oriental Region. Molecular Ecology 23: 2573–2589.

Zhang D-X, Hewitt GM, 2003. Nuclear DNA analyses in genetic studies of populations: Practice, problems and prospects. Molecular Ecology 12: 563–584.

Supplementary Material

Table S1 Dopamine receptor D4 (*DRD4*) gene alleles from a 555 bp segment including partial exon 3 and intron 3

	Exon 3						Intron 3						
Allele	26	56	92	125	131	248	282	300	369	411	427	481	542
1	G	G	T	G	C	T	G	C	A	G	G	G	G
2	G	G	T	G	T	T	G	C	G	G	C	G	G
3	C	G	T	G	T	T	T	C	G	C	C	G	G
4	C	G	T	G	T	T	G	T	A	G	G	G	G
5	C	G	T	G	T	T	G	C	A	G	G	G	G
6	G	G	T	G	C	T	G	C	G	G	G	G	A
7	G	G	T	G	C	T	G	T	A	G	G	G	A
8	C	G	T	G	T	T	G	T	A	G	G	G	A
9	C	G	C	G	T	T	G	T	A	G	G	G	A
10	C	G	C	G	T	T	G	C	G	G	C	T	G
11	C	G	T	G	T	T	G	T	A	G	C	G	A
12	G	G	T	G	T	T	G	C	A	G	C	G	G
13	C	G	C	G	T	T	G	C	A	G	G	G	G
14	G	G	T	G	T	T	G	C	G	G	G	G	A
15	C	G	T	A	T	T	G	C	G	G	G	G	A
16	C	G	C	G	T	T	G	C	G	G	C	G	G
17	C	G	C	G	T	T	G	T	A	G	G	G	G
18	C	G	T	G	T	T	T	C	A	C	C	G	G
19	C	G	T	A	T	T	G	C	G	G	C	G	G
20	C	G	T	G	T	C	G	C	A	G	G	G	G
21	G	G	C	G	T	T	G	C	A	G	G	G	G
22	C	G	T	G	T	T	G	C	G	G	C	G	G
23	G	G	T	G	C	T	G	C	G	G	C	T	G
24	C	G	T	G	T	C	G	T	A	G	G	G	G
25	G	C	T	G	T	T	G	C	G	G	C	G	G
26	C	G	C	G	T	T	G	T	A	G	C	G	A
27	C	G	C	G	T	T	G	C	G	G	C	G	A
28	C	G	T	G	C	T	G	T	A	G	G	G	A
29	G	G	T	G	T	T	G	C	G	G	C	T	A
30	G	G	T	G	T	T	G	C	G	G	C	T	G
31	C	G	T	G	T	T	G	C	A	G	C	G	G
32	C	G	C	G	T	T	G	C	A	G	C	T	G
33	C	G	T	G	T	C	G	T	A	G	G	G	A
34	C	G	C	G	T	T	T	C	G	C	C	G	G
35	G	C	T	G	T	T	G	C	G	G	G	G	A
36	C	G	T	G	C	T	G	C	G	G	C	G	G
37	C	G	T	G	T	T	G	C	G	G	G	G	A
38	G	G	T	G	T	T	T	C	G	C	C	G	G
test													

The nucleotide at each polymorphic site for each allele is shown. Full sequence deposited in GenBank (KP941028).

Table S2 ***DRD4*** **exact test pairwise *p*-values (significant results after FDR correction denoted in bold)**

	WM/J	WMN	WCB	WCO	SNU	SCR	SST	STB	SAS	SMA	SMC	SMN	VYV	NOR	TN/S
WM/J	—														
WMN	0.851	—													
WCB	0.187	0.763	—												
WCO	**0.000**	**0.000**	**0.000**	—											
SNU	**0.000**	**0.000**	**0.000**	**0.000**	—										
SCR	**0.000**	**0.000**	**0.000**	**0.000**	0.587	—									
SST	**0.000**	**0.000**	**0.000**	**0.000**	0.327	0.643	—								
STB	**0.000**	**0.000**	**0.000**	**0.000**	0.486	0.263	0.042	—							
SAS	**0.000**	**0.000**	**0.000**	**0.000**	0.060	0.241	**0.023**	0.118	—						
SMA	**0.000**	**0.000**	**0.000**	**0.000**	**0.021**	0.889	0.543	**0.002**	**0.007**	—					
SMC	**0.000**	**0.000**	**0.000**	**0.000**	**0.033**	0.886	0.173	**0.012**	0.041	0.844	—				
SMN	**0.000**	**0.000**	**0.000**	**0.000**	**0.001**	0.534	0.219	**0.006**	0.035	0.419	0.460	—			
VYV	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	—		
NOR	**0.000**	**0.000**	**0.000**	**0.000**	**0.009**	0.102	**0.033**	**0.025**	**0.009**	0.168	0.039	**0.010**	**0.001**	—	
TN/S	**0.000**	**0.000**	**0.000**	**0.000**	0.168	0.302	0.618	0.138	0.156	0.070	0.123	0.346	**0.000**	**0.002**	—