CrossMark

ORIGINAL PAPER

The genetic structure of the introduced house sparrow populations in Australia and New Zealand is consistent with historical descriptions of multiple introductions to each country

Samuel C. Andrew [ORCID] **· Monica Awasthy · Peri E. Bolton · Lee A. Rollins ·
Shinichi Nakagawa · Simon C. Griffith**

Abstract The house sparrow is one of the most widely introduced vertebrate species around the world, making it an important model species for the study of invasion ecology. Population genetic studies of these invasions provide important insights into colonisation processes and adaptive responses occurring during invasion. Here we use microsatellite data to infer the population structure and invasion history of the introduced house sparrow (*Passer domesticus*) in Australia and New Zealand. Our results identify stronger population structure within Australia in comparison to New Zealand and patterns are consistent with historical records of multiple introduction sites across both countries. Within the five population clusters identified in Australia, we find declines in genetic diversity as we move away from the reported introduction site within each cluster. This pattern is consistent with sequential founder events. Interestingly, an even stronger decline in genetic diversity is seen across Australia as we move away from the Melbourne introduction site; secondary historical reports suggest this site imported a large number of sparrows and was possible the source of a single range expansion across Australia. However, private allele numbers are highest in the north, away from Melbourne, which could be a result of drift increasing the frequency of rare alleles in areas of smaller population size or due to an independent introduction that seeded or augmented the northern population. This study highlights the difficulties of elucidating population dynamics in introduced species with complex introduction histories and suggests that a combination of historical and genetic data can be useful.

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s10530-017-1643-6) contains supplementary material, which is available to authorized users.

S. C. Andrew (✉) · M. Awasthy ·
P. E. Bolton · S. C. Griffith
Department of Biological Sciences, Macquarie University, Sydney, NSW 2109, Australia
e-mail: samuel.andrew@students.mq.edu.au

L. A. Rollins
Centre for Integrative Ecology, School of Life and Environmental Sciences, Deakin University, Pigdons Road, Geelong, VIC 3217, Australia

S. Nakagawa
Department of Zoology, University of Otago, Dunedin 9054, New Zealand

S. Nakagawa
Evolution and Ecology Research Centre and School of Biological, Earth and Environmental Sciences, The University of New South Wales, Sydney, NSW 2052, Australia

Introduction

The recent and rapid colonization history of invasive species can make them valuable models for the study of colonization events and their ecological impacts (Barrett 2014). Exotic introductions also provide excellent 'natural' experiments to test hypotheses about local adaptation in species that have been introduced to different environmental conditions than are found in their native range (Lee 2002; Sax et al. 2007; Blackburn 2008). Studies of colonization ecology or the extent to which introduced species have become locally adapted benefit from background knowledge such as the time of arrival and the identity of the original source population. To an extent this information can be reconstructed or confirmed using population genetic approaches (Estoup and Guillemaud 2010). Introduced and/or invasive populations are regularly affected by multiple introductions, population bottlenecks and successive founder effects as they colonize new areas (Dlugosch and Parker 2008). These colonization processes leave genetic signatures that can be used to reconstruct invasion histories (Benazzo et al. 2015; Cristescu 2015). A challenge for studies of local adaptation is that these 'genetic signatures' of colonization history and population structure can hamper the use of genetic methods to identify true signals of local adaptation (Excoffier et al. 2009b; Günther and Coop 2013; de Villemereuil et al. 2014; Lotterhos and Whitlock 2014). As a result, population genetic data are valuable in helping us to characterize and limit the confounding effect of these phenomena, before the successful implementation of genomic studies of local adaptation (De Mita et al. 2013; Rellstab et al. 2015; Francois et al. 2016).

Without the benefit of population genetic data, invasion history may be based on assumptions derived from historical records. These are inherently prone to error because their sources may be unreliable or may lack a global understanding of the invasion. Potential complexities that can result in inaccurate accounts include multiple unreported introductions to the same, or different localities. Due to direct and indirect anthropogenic influences, the colonization history of invasive species is likely to be more complex than naturally distributed species (Miller et al. 2005). In many cases these anthropogenic colonization events have even been organised and intensive, with some of

the most notable examples of introductions being from global chapters of Acclimatisation Society in the mid 1800s (Lever 1985, 1992, 2005). At the time when Acclimatisation Societies were active, private introductions were also taking place, that were motivated by similar philosophies, but were probably less well documented (Balmford 1981).

Acclimatisation Societies focused on a number of species but a few, including the house sparrow (*Passer domesticus*), were introduced widely across the world. Currently, the house sparrow is one of the most broadly distributed bird species in the world (Anderson 2006), largely due to human introductions starting in the mid 1800s to North America, South America, Australasia and Africa (Lever 2005). The species' native distribution covers most of the Palaearctic; this distribution was probably established after the species formed a commensal relationship with humans about 10,000 years ago and spread throughout Eurasia, concurrent with the spread of agriculture from the Middle East (Sætre et al. 2012). The house sparrow was introduced to Australia and New Zealand in the 1860s mostly by Acclimatisation Societies (Lever 2005). The species was a very successful colonizer and expanded its distribution to cover almost all the climatic conditions across Australasia (Higgins et al. 2006). Although Acclimatisation Societies kept detailed records of their introductions, many details may have been lost or miscommunicated through time. This is demonstrated by an investigation of primary literature for house sparrow introductions to Australia, which uncovered repeated introductions, new successful introductions and source populations that were previously unrecognised in the scientific literature (Andrew and Griffith 2016). The house sparrow was also introduced to New Zealand from England in the 1860s with five reported introduction points [see Table S1 and Thomson (1922)].

The house sparrow's broad distribution and its close proximity to humans has made this species an excellent and broadly studied model species for invasion genetics (Liebl et al. 2015). However, to date the introduction history and local adaptation in Australasian house sparrow populations has not been studied using modern molecular techniques. Early work on Australian and New Zealand populations using allozymes found that introduction events reduced genetic diversity and increased differentiation in introduced populations compared to native

populations in Europe (Parkin and Cole 1985; Cole and Parkin 1986). More recent molecular techniques have been applied to native and introduced house sparrow populations around the world. These studies cover a number of topics including population structure (Schrey et al. 2011, 2014; Kekkonen et al. 2011b; Jensen et al. 2013) demographic factors (Vangestel et al. 2011; Kekkonen et al. 2011a; Baalsrud et al. 2014) and observing the link between phenotypic and genetic differentiation (Lima et al. 2012). The Australasian house sparrow populations provide a nice opportunity to replicate findings from other introductions and examine the evolution of the species in the context of the Australasian landscape and climate.

Genetic population structure and demographic history are unique for each introduction and should be characterised as part of the basic biology of an introduced species. Here we investigated the introduced house sparrow populations of both Australia and New Zealand. We predict: (1) independent introduction events with varied propagule size and origins will have caused genetic differentiation and population structure between introduction sites due to initial differences in allele frequency and genetic diversity; (2) range expansions from the original sites of introduction will have caused declines in genetic diversity with successive founder events (Peter and Slatkin 2015); (3) successive founder events will have resulted in population differentiation that is strongest at the range edge, due to genetic drift; (4) Across the broad geographic sampling range in Australia we expect to see a pattern like isolation by distance (IBD) due to the recent colonization and the species low natural dispersal ability across the large distances of uninhabitable habitat between isolated human settlements (in this highly commensal species). We discuss the relevance of our findings to historical records of the introduction of these house sparrow populations and related systems.

Materials and methods

Sampling

Adult house sparrows were collected from 25 urban localities across Australia with approximately 40 birds sampled at each locality (total number of birds genotyped = 1027, Table 1, Fig. 1). The Australian samples were collected during April to September 2014 and March 2015 under the Animal Research Authority of the Animal Ethics Committee at Macquarie University (ARA 2014/248). Samples from New Zealand were collected from four localities between June and August 2005, under the ethics approval of Otago University (Animal ethics reference number: 87/08). Approximately 40 individuals were genotyped from each of the four localities in New Zealand (n = 170, Table 1). Blood samples from a population in Morthen, South Yorkshire, UK were also sourced from a previous study to be included as a comparison for an English source population (n = 40) (Ockendon et al. 2009). Birds were captured using mist nets and placed in bird bags until a blood sample could be taken. Blood was taken from the brachial vein with a capillary tube (ca. 40 μl) and was stored in 800 μl of absolute ethanol in a 1.5 ml microcentrifuge tube. Birds were not held for more than 30 min and were released as soon as they had been sampled and banded. In total, 1237 birds were genotyped from 30 localities (Table 1, Fig. 1).

Molecular methods

DNA was extracted using a Gentra PureGene tissue kit (Qiagen, Valencia, CA, USA) following manufacturer's instructions. Samples were genotyped using two multiplexes developed by Dawson et al. (2012) which included 13 polymorphic loci and a sexing locus (Multiplex 1: Ase18, Pdoμ1, Pdoμ3, Pdoμ5, Pdoμ6, Pdo9, Pdo10, P2D/P8; Multiplex 2: Pdo16A, Pdo17, Pdo19, Pdo22, Pdo27, Pdo40A). PCRs were carried out using 5 μl reactions. For each reaction 1 μl of genomic DNA (ca. 100 ng/μl) was added to 2.5 μl of Master Mix (Qiagen, Valencia, CA, USA), 0.5 μl of primer mix [see concentrations in Dawson et al. (2012)] and 1 μl of Milli-Q water. Both multiplex reactions used the same PCR thermal cycle with a hot-start denaturing phase of 10 min at 95 °C followed by 33 cycles of 94 °C for 30 s, 57 °C for 90 s and 72 °C for 90 s, before a final extension at 72 °C for 10 min. The post PCR product was diluted and genotyped on the ABI 3730XL DNA analyzer using GS500 (LIZ) as a size standard for Multiplex 1 and GS1200 for Multiplex 2. Microsatellite alleles were scored using the GeneMapper program version 3.7 (data included with online supplementary material).

Table 1 Population information including: locality ID number (no.); location name; state within country; name of population cluster (population cluster); sample size (*n*); the year the population was reported as established (year est.); distance from the proposed introduction site [Dist. (km)]; mean allelic richness across loci (mean Ar); mean observed heterozygosity (mean H_O); mean expected heterozygosity (mean H_E)

No.	Location	State	Population cluster	*n*	Lat	Long	Year est.	Dist. (km)	Mean Ar	Mean H_O	Mean H_E
1	Tennant Ck	NT	Brisbane	40	− 19.657	134.192	1978	1986	6.636	0.698	0.679
2	Mt Isa	QLD	Brisbane	77	− 20.730	139.503	1965	1454	10.248	0.714	0.739
3	Longreach	QLD	Brisbane	43	− 23.436	144.255	1970	880	9.308	0.729	0.722
4	Atherton	QLD	Brisbane	42	− 17.215	145.480	1965	1328	9.390	0.725	0.728
5	Townsville	QLD	Brisbane	42	− 19.319	146.824	1965	1055	9.779	0.738	0.725
6	Charleville	QLD	Brisbane	43	− 26.403	146.251	1960	553	9.545	0.704	0.715
7	Roma	QLD	Brisbane	42	− 26.562	148.791	1955	309	9.281	0.686	0.713
8	Toowoomba	QLD	Brisbane	42	− 27.722	151.632	1875	0	10.833	0.735	0.774
9	Armidale	NSW	Sydney	44	− 30.516	151.674	1905	354	11.273	0.812	0.816
10	Dubbo	NSW	Sydney	39	− 32.221	148.626	1905	257	10.628	0.790	0.791
11	Sydney	NSW	Sydney	42	− 33.620	150.816	1865	0	10.316	0.798	0.801
12	Goulburn	NSW	Sydney	44	− 34.763	149.699	1905	163	12.188	0.783	0.778
13	Cobar	NSW	Syd/Melb	38	− 31.492	145.830	1905	524	12.112	0.805	0.819
14	Wentworth	NSW	Melbourne	39	− 34.105	141.917	1905	490	13.624	0.809	0.826
15	Burrumbuttock	NSW	Melbourne	25	− 35.837	146.802	1905	274	11.199	0.813	0.831
16	Albury	NSW	Melbourne	35	− 36.066	146.936	1905	274	12.364	0.855	0.829
17	Melbourne	VIC	Melbourne	42	− 37.789	144.915	1863	0	12.936	0.807	0.819
18	Geelong	VIC	Melbourne	40	− 38.178	144.372	1863	64	14.273	0.823	0.851
19	Mt Gambier	SA	Adel/Melb	41	− 37.857	140.848	1868	360	11.485	0.779	0.790
20	Broken Hill	NSW	Adelaide	41	− 31.946	141.465	1897	458	11.836	0.810	0.810
21	Adelaide	SA	Adelaide	42	− 35.232	138.489	1865	0	10.974	0.776	0.809
22	Coober Pedy	SA	Adelaide	20	− 29.006	134.748	2000	777	7.694	0.818	0.765
23	Wynyard	TAS	Hobart	38	− 40.972	145.653	1875	260	12.363	0.832	0.821
24	Bridport	TAS	Hobart	43	− 41.000	147.388	1875	215	11.038	0.826	0.819
25	Hobart	TAS	Hobart	43	− 42.936	147.351	1867	0	11.302	0.765	0.797
26	Auckland	North Is	North NZ	41	− 36.865	174.776	1867	0	13.182	0.785	0.805
27	Wellington	North Is	North NZ	43	− 40.623	175.288	1866	0	13.364	0.791	0.820
28	Christchurch	South Is	South NZ	42	− 43.532	172.627	1867	0	11.727	0.836	0.833
29	Dunedin	South Is	South NZ	44	− 45.876	170.496	1868	0	13.273	0.829	0.839
30	Morthen	UK	South NZ	40	51.75	− 1.253	NA	NA	12.975	0.813	0.849

Genetic analyses

Missing data percentages were calculated in Microsoft Excel, and loci with more that 5% missing data were excluded. Deviations from Hardy–Weinberg equilibrium (HWE) were tested in ARLEQUIN version 3.5.2.2 (Excoffier et al. 2005). Null allele frequency estimates were calculated using CERVUS (Marshall et al. 1997) and loci with null allele frequencies greater than 10% were excluded. GENEPOP (Raymond and Rousset 1993) was used to test linkage disequilibrium (LD) between loci within sampling localities. Allelic
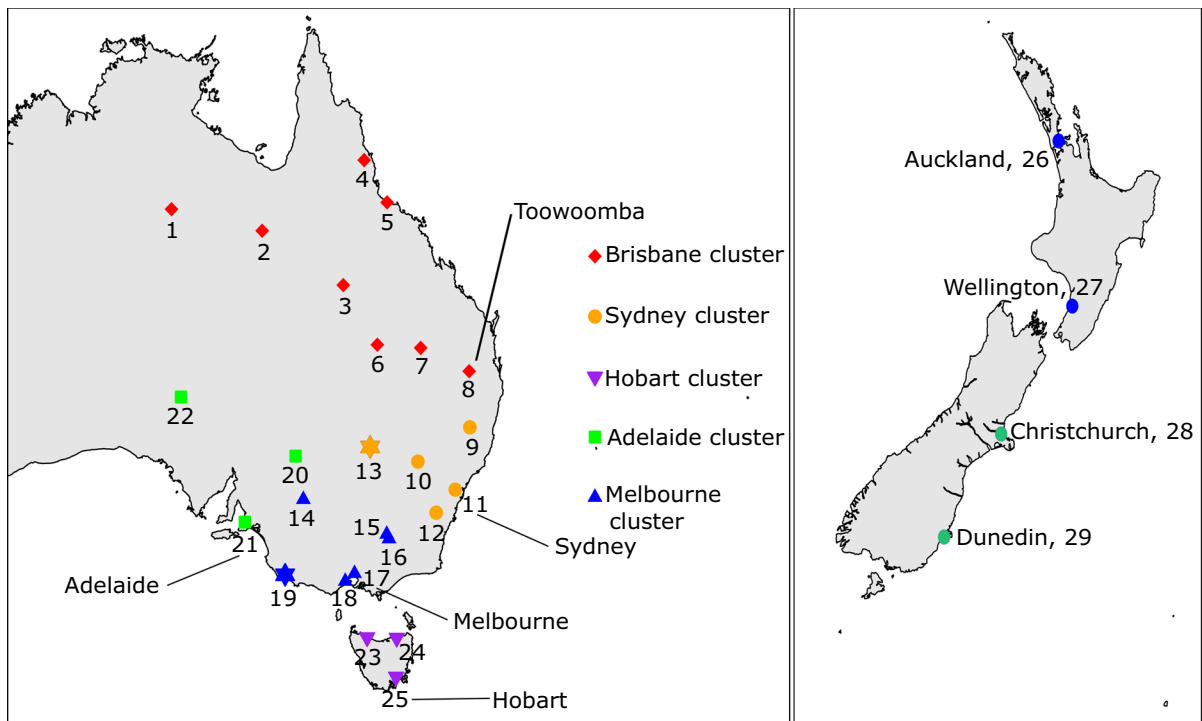
Fig. 1 Map of sampling sites. The numbers next to the points for the sampling localities are the same as the ID numbers in Table 1. The colour coding is linked to the genetic population clusters described in Fig. 2. The five labelled sites on Australia are those closest to the major cites that housed acclimatisation society chapters that are link to the sparrows' introduction. The two stars represent two sample sites with uncertain population allocations possible due to admixture

richness, number of alleles and genetic diversity (expected heterozygosity) were calculated using FSTAT version 2.9 (Goudet 1995). ARLEQUIN was used to produce a pairwise F_{ST} matrix and to run Analysis of molecular variance (AMOVA) to estimate among population differentiation (F_{ST}), for all sample sites and for pooled Australian and pooled New Zealand samples separately. GENALEX version 6.503 (Peakall and Smouse 2006, 2012) was used to calculate private allele frequencies. Using R version 3.3.1 (R Core Development Team 2017), allele frequency heat maps were drawn for each locus to show differences between population clusters, as a graphical aid to describe population genetic diversity.

STRUCTURE (Pritchard et al. 2000) was used to identify populations using a model-based Bayesian clustering method that calculates the probability an individual belongs to a cluster when a given number of clusters (K) is specified. To infer the number of clusters with the most support in each analysis we used the delta K method (Evanno et al. 2005). Delta K and mean LnP(K) and other summary statistics were calculated in STRUCTURE HARVESTER (Earl and VonHoldt 2012). The delta K method will identify the upper most level of structure, we also looked at hierarchical population structure within the genetic groups defined by our initial models (Rollins et al. 2009). For our final STRUCTURE analyses we used a MCMC length of 1,000,000 iterations and a burn-in period of 100,000 and 10 repeats for each value of K using an admixture model. The range of K values used in our final runs were chosen based on shorter preliminary runs with fewer iterations (200,000–400,000) that included values of K equal to our number of sample localities. We confirmed the peak in delta K was a true signal by checking that the variance in mean LnP(K) was stable between values of K because some repeats were not converging (see Table S2). However, the delta K method will not show if the greatest support is for one cluster. Therefore, for all models we determined if the greatest support was for one cluster (K = 1) by checking if mean LnP(K) was highest for K = 1. Q-plots for the most highly supported value of K were drawn using the

results of all 10 repeats with the programs CLUMPP (Jakobsson and Rosenberg 2007) and DISTRUCT (Rosenberg 2004).

The STRUCTURE method relies on demographic assumptions about the study populations that are rarely met in the real world (e.g. no pattern of IBD). For this reason, we also used a second method to look at population structure that does not make assumptions about demographic models. The R package *adegenet* (Jombart 2008) was used for a Correspondence Analysis (CoA) of microsatellite data for all 30 localities; this multivariate approach uses a summary of sample site allele frequencies to create a distance matrix that is used to generate Principal Component (PC) values for each locality, similar to a Principal Coordinate Analysis (PCoA) of individuals. Our CoA used 5 PC's because we only had 30 sample localities (PC's must be less than *n*), and were enough to explain almost all the variance in the data. To visualise patterns of population structure which can be compared with the results from the STRUCTURE clustering analyses, we used the PC values from the CoA in a Discriminant Analysis of Principal Components (DAPC) in *adegenet* (Jombart et al. 2009, 2010). The number of genetic clusters was inferred using the "find.clusters" function and the optimal number of clusters was decided based on BIC reduction. The cluster labels were then used in the first DAPC of the PCs from the CoA to make a scatter plot. Using the cluster labels defined by the "find.clusters" function, we ran a second DAPC using the individual data to calculate the percentage of individuals that were correctly assigned to their population clusters defined by the first DAPC of localities. We choose to use the PC from a CoA of sample site (that uses allele frequencies, also accounting for the presence and absence of alleles) to define the main genetic clusters, because we predict founder effects will have had the clearest effect on allele frequency and allelic diversity between localities.

Neighbour-joining trees are often used to reconstruct invasion histories by using the branching pattern to describe invasion routes (Estoup and Guillemaud 2010). We tested if there were any clear invasion routs across Australia by drawing a neighbour-joining tree based on the genetic distance metric of Cavalli-Sforza and Edwards (1967) with the program POPULATIONS (Langella 2002). Bootstrap values for the tree

were calculated over 1000 iterations that used different subsets of loci.

To test for isolation by distance (IBD) we used Mantel tests, in R using the package *adegenet* (Jombart 2008). We used the function "mantel.randtest" which used 999 replications for the tests. The Mantel tests used Edwards' Euclidean distance (2nd method option) to calculate genetic distance. We ran this analysis for all Australian sampling sites and for the larger population clusters independently to look for differences in connectivity at different special scales. We also modelled the effects of sequential founder effects on genetic differentiation (pairwise F_{ST}) within population clusters. For this question, the response variable was the F_{ST} between each sample site and the sample site nearest the original historic introduction site (this removed introduction sites from the model). The predictor was "year established" (see methods below) and we expected F_{ST} to be higher for more recently established populations.

To infer the relationship between colonization history and genetic diversity we used Linear Mixed Models (LMM). The LMMs were run in R using the package *lme4* (Bates et al. 2015) with the package *lmerTest* (Kuznetsova et al. 2016) to calculate degrees of freedom and *P* values. To model the effects of range expansion and founder events on genetic diversity we collected data on the "year established" (the year populations were first recorded as being present in a locality, as previously described in Andrew and Griffith (2016)] and the "distance from source" (distance from the proposed original introduction site location (km)) to be used as fixed effects. The two fixed effects of "year established" and "distance from source" were found to be highly correlated (estimate = 0.072, $t_{27} = 7.664$, $P < 0.0001$, $R^2 = 0.685$), so could not be used together in the same model. We chose to use year established as a proxy for range expansion and sequential founder events in the LMMs. Demographic and genetic non-independence was accounted for by using the "population" clusters from the DAPC analysis as a random factor. The response variables used to measure genetic diversity were allelic richness and expected heterozygosity; these response variables were used in separate models with the same structure as described above. We calculated marginal R^2 and intra-class correlation coefficients (ICC) values for all LMM's using the method described in Nakagawa and Schielzeth (2013). To

visualise the patterns in the data we used scatter plots for all combinations of the two response variables vs the two predictor variables, and a third predictor variable "Sample site distance from Melbourne" was also plotted. Lines of best fit were plotted using a linear regression between the two variables.

Results

We used 11 polymorphic loci (Ase18, Pdoμ1, Pdoμ3, Pdoμ6, Pdo10, Pdo16A, Pdo17, Pdo19, Pdo22, Pdo27, Pdo40A) for analyses after removing Pdo9 due to more than 5% missing data. Pdoμ5 was also removed for being out of HWE in more than 20% of localities, which may be owing to a high null allele frequency (15.1%, Table S3). The remaining loci were not out of HWE in more than 5% of the sample sites. No loci were found to be consistently affected by LD within sampling localities. In total, we genotyped 1237 individuals from 30 sample sites, summary genetic diversity statistics for each sampling locality are presented in Table 1.

Tests of genetic differentiation using an AMOVA found significant differentiation among sampling localities. For all 30 localities, among population $F_{ST} = 5.60\%$ ($df = 29$, $P < 0.001$); for only Australian samples, among populations $F_{ST} = 6.01\%$ ($df = 24$, $P < 0.001$); and for only New Zealand samples, among population $F_{ST} = 1.90\%$ ($df = 3$, $P < 0.001$, see Table S4 for all AMOVA results). Pairwise F_{ST} comparisons also found strong evidence for genetic differentiation between sampling localities, with significant differentiation in over 95% of pair-wise comparisons (420 out of 435) after Bonferroni corrections (Fig. S1).
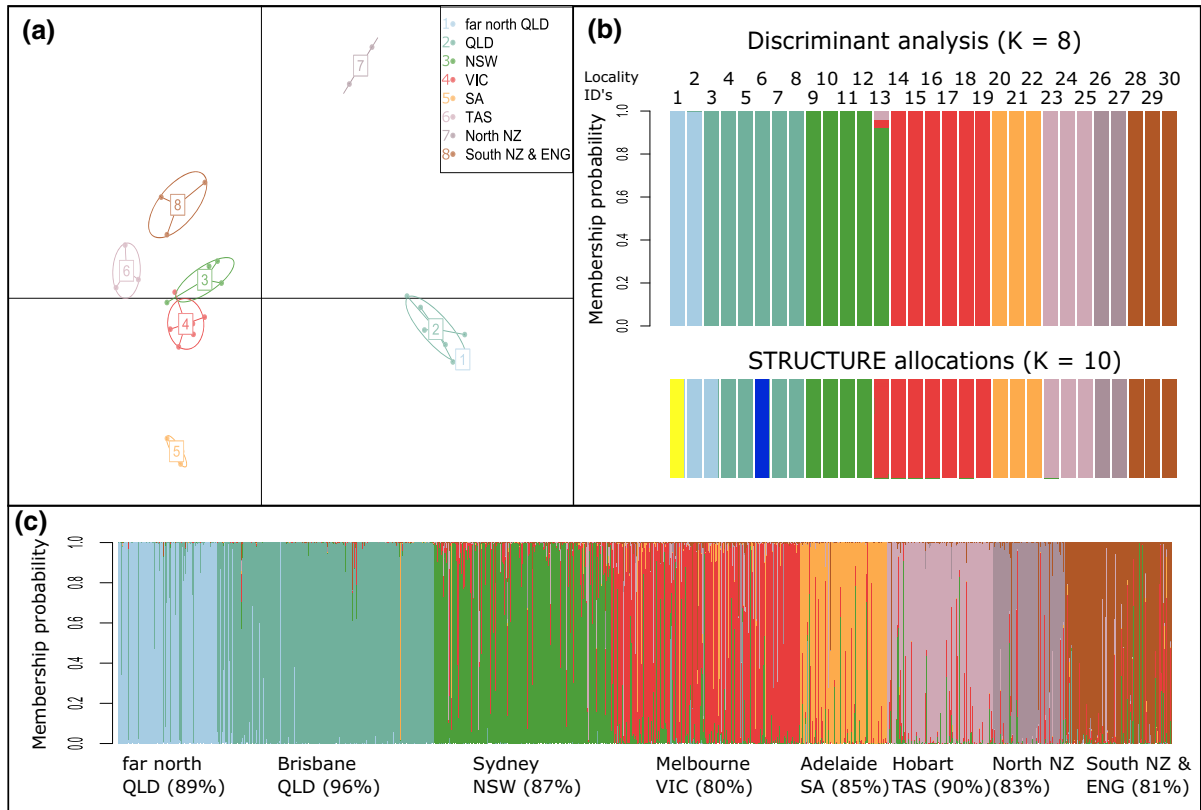
Fig. 2 Discriminant analysis of principle components. **a** Shows the scatter plot for the DAPC of the 30 sample localities that found 8 clusters (see figure key). **b** Shows the membership probability of each locality to the clusters, only 13 (Cobar) had mixed membership. Below the membership probabilities is a visual summary of how localities were allocated to clusters using STRUCTURE (see Fig. S2 for Q-plots). **c** Uses the 8 genetic population ID's from **a** and calculates the membership probabilities of individuals using a second DAPC. The percentages show the proportion of individuals correctly assigned to their predicted cluster

Population structure was visualised using a DAPC approach (Fig. 2). A multivariate analyses (CoA) of population structure that used a distance matrix for the 30 sample sites based on allele frequencies, was used in an analysis to identify the number of meaningful clusters in the data that provided the lowest BIC. We found BIC levelled off and stopped improving at eight clusters (Fig. 2a, b). The eight cluster ID's for the sample sites were then applied to the individual genotype data and used to run a second DAPC. We found that greater than 80% of individuals (range 80–96%) were correctly assigned to their population ID's from the first DAPC, giving support for these population clusters (Fig. 2c). The names for the eight clusters in Fig. 2 are based on the capital city within a polygon drawn around the clusters sample sites or the geographic region. Apart from these eight clusters the DAPC also visualises a very clear divide between the "northern Australia" localities and all the other localities (including southern Australia, New Zealand and England).

The results from the model-based clustering analysis in STRUCTURE (Fig. S2) were similar to the results from the DAPC approach. After accounting for substructure we found 10 clusters using STRUCTURE which are summarised in Fig. 2b. The only inconsistencies between the STRUCTURE and the DAPC method was that the northern Australia localities are broken into four rather than two clusters respectively and STRUCTURE grouped Cobar with the Melbourne cluster rather than with the Sydney cluster (Fig. 2b). The scatter plot in Fig. 2a, however, shows that the eight northern Australia localities (hence referred to as Brisbane cluster) are closely grouped, indicating weak structure within that region; The Melbourne and Sydney clusters were also relatively close together. A neighbour-joining tree of Australian localities found localities from the same cluster were grouped within clades but the branching pattern of the tree did not have any significant support (most boot strap values < 70%) for possible invasion roots across Australia (Fig. 3).

Population structure established by founder effects could have been eroded since the original colonization events due to gene flow. If gene flow is low then regions that had independent shipments of sparrows into Australia could have maintained private alleles. We visualise allele presence absence between genetic populations using heat maps (Fig. S3). Predictably,

New Zealand and England had alleles that were not observed in Australia for most loci. The native English sample site also had relatively high genetic diversity when compared to the invasive populations (Table 1). Between the five Australian population clusters, the total number of private alleles (including those with minor frequencies) were: Melbourne = 9, Brisbane = 8, Sydney = 4, Adelaide = 4, Hobart = 1. The number of private alleles with a frequency > 1% (to avoid falsely identifying rare alleles that are easily missed without comprehensive sampling) are much lower, the Brisbane cluster had three, the Melbourne cluster had two and the Hobart cluster had one private allele.

To further look at patterns of dispersal, we tested for IBD. The Mantel test found strong support for IBD across sample sites within Australia ($R^2 = 0.758$, n = 25, P = 0.001). This pattern is consistent within the two main population clusters of Brisbane (northern Australia, $R^2 = 0.545$, n = 8, P = 0.024), southern Australia ($R^2 = 0.767$, n = 17, P = 0.001) and for the Melbourne cluster within southern Australia ($R^2 = 0.520$, n = 6, P = 0.020) when analysed separately (see Fig. S4 for details). Other sub-clusters had a smaller number of sample localities so were not also analysed separately. However, the relationships were less strong within the Brisbane and Melbourne clusters (Fig. S4).

In invasions, sequential founder events can reduce genetic diversity and increase F_{ST}. We predicted that sequential founder effects have created population differentiation, where pairwise F_{ST} between sample sites (and the puta tive source) should be positively correlated with the "year established". Using linear models, we found a significant positive relationship between the "year established" and the "pairwise F_{ST} comparison with the sample site nearest the putative site of introduction" (estimate = 0.0003, $t_{23} = 4.651$, P < 0.001, $R^2 = 0.485$).

The LMMs for the effects of range expansion (year established) on genetic diversity also found allelic richness and expected heterozygosity declined in more recently established populations (Table 2). We also found that the random factor of "population" explained a large proportion of the variance in the data (Table 2), where lower intercepts corresponded to lower levels of genetic diversity (Fig. S5). We have used "year established" as a proxy measure of sequential founder events, but distance from the
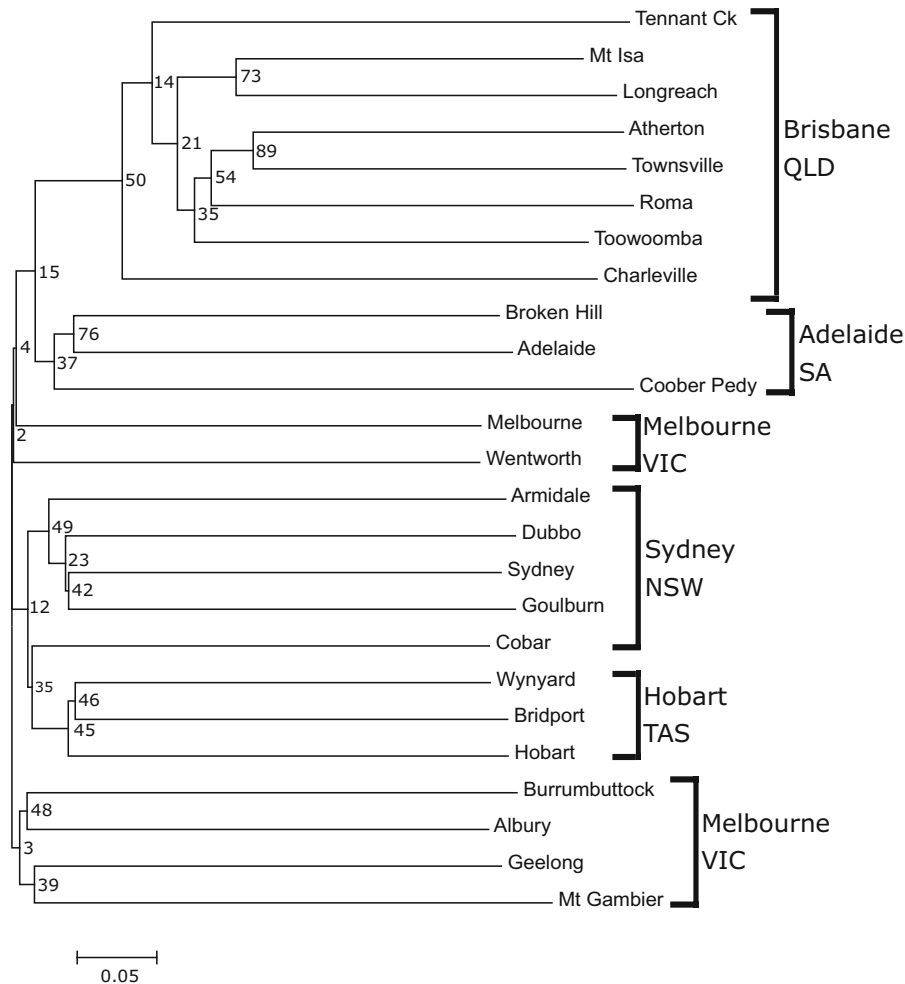
Fig. 3 Neighbour joining tree of Australian localities. Labels to the right link localities in the same population cluster from the DAPC (Fig. 2). The Tree puts the sample sites linked to the Brisbane cluster in a clade with a large separation from the other sites. Sample sites linked to the Sydney, Adelaide and Hobart clusters are consistently grouping in their own independent clades. Cobar is on a branch between the Sydney and Hobart clades showing again that it is not consistently grouping with the same localities in different analyses. The localities linked to the Melbourne population are mixed across multiple clades. Bootstrap percentages for nodes are also included on the tree, calculated using subsets of loci. In general sister branches on the tree are sample sites with low pair-wise F_{ST} (Fig. S1). Bootstrap values are relatively high within clades but there is no clear relationship between clades illustrating invasion roots across Australia

putative introduction site is also correlated with this variable and genetic diversity (Fig. 4). The distance from the putative introduction site was negatively correlated with genetic diversity (Fig. 4b, e). The negative correlation with genetic diversity is even stronger when we calculate the distance of each site from Melbourne (Fig. 4c, f). However, Melbourne is in the south of Australia and genetic diversity is higher in the south versus the north, allelic richness ($t_{23} = 4.506$, $P < 0.001$, $R^2 = 0.469$) and expected hetrozygosity ($t_{23} = 7.007$, $P < 0.0001$, $R^2 = 0.681$)

are both positively correlated with latitude. Therefore, this relationship for Melbourne could be true for any southern locality. A summary for the linear models for the lines of best fit in these scatter plots in Fig. 4 are given in Table S5.

Discussion

Across our Australian and New Zealand sampling locations and the founding population (England), we

Table 2 Effects of range expansion on allelic richness and genetic diversity (H$_E$)

|	Estimate	SE	df	t	P
(a) Mean allelic richness					
Intercept	64.32	10.61	16.1	6.06	< 0.0001
Year established	− 0.028	0.006	15.9	− 5.00	< 0.0001
	Variance	SD	n		ICC
Marginal R^2	1.473				0.527
Population	0.227	0.477	6		0.081
Residual	1.095	1.047	29	Conditional R^2	0.608
	Estimate	SE	df	t	P
(b) Mean genetic diversity					
Intercept	1.667	0.23	26.4	7.33	< 0.0001
Year established	− 0.0004	0.0001	26.4	− 3.84	0.0007
	Variance	SD	n		ICC
Marginal R^2	0.0004				0.301
Population	0.0006	0.024	6		0.437
Residual	0.0003	0.019	29	Conditional R^2	0.739

Year of population establishment and distance from source introduction could not be used as fixed effects in the same model due to high collinearity. We choose to use "year established" as a proxy for sequential colonization events. There is a consistent significant negative relationship between the year populations were established and genetic diversity. These relationships are plotted in Fig. 4. Mean allelic richness and expected heterozygosity was calculated using the 11 loci for the 29 populations in Australian and New Zealand. The marginal R^2 is the variance explained by the fixed effect of year established and the Conditional R^2 is for the total model

identified significant population structuring with eight population clusters identified by the DAPC method (Fig. 2). The main population structure in Australia was found between northern Australia (Brisbane cluster) and the rest of the sampling localities (southern Australia). The two population 'sub-clusters' within the Brisbane cluster were very genetically similar and this separation is likely due to sequential colonization events and isolation. Therefore, we refer to the eight northern localities as the Brisbane cluster. The English locality clustered with the South Island of New Zealand although they are clearly independent reproductive populations but have similar genetic compositions. This population structure was also supported using a Bayesian clustering analysis (Fig. S2) which suggests these results are repeatable and relatively robust.

We find evidence of IBD across our Australian sampling localities, this violates one of the assumptions of the widely used STRUCTURE analyses and could potentially influence our results (Frantz et al.

2009; Jombart et al. 2009). Therefore, we have focused on a DAPC method to identify population structure. We suggest the DAPC method will be useful in future studies of population structure in invasive species with complex introduction histories because the method does not rely on demographic assumptions (Jombart et al. 2009, 2010). This method also has flexibility in its application because the genetic data can be used to define population clusters or to test the accuracy of predefined clusters. Predefined populations could be based on sampling design, the species biology or results from related analyses, as was done here in the second DAPC.

The high level of genetic differentiation found among populations (AMOVA and pairwise F$_{ST}$) across both Australia and New Zealand is expected for a relatively sedentary species such as the house sparrow that has gone through sequential colonization events. Levels of population differentiation (pairwise F$_{ST}$) reported for house sparrow populations around the world report some similar patterns (Schrey et al.
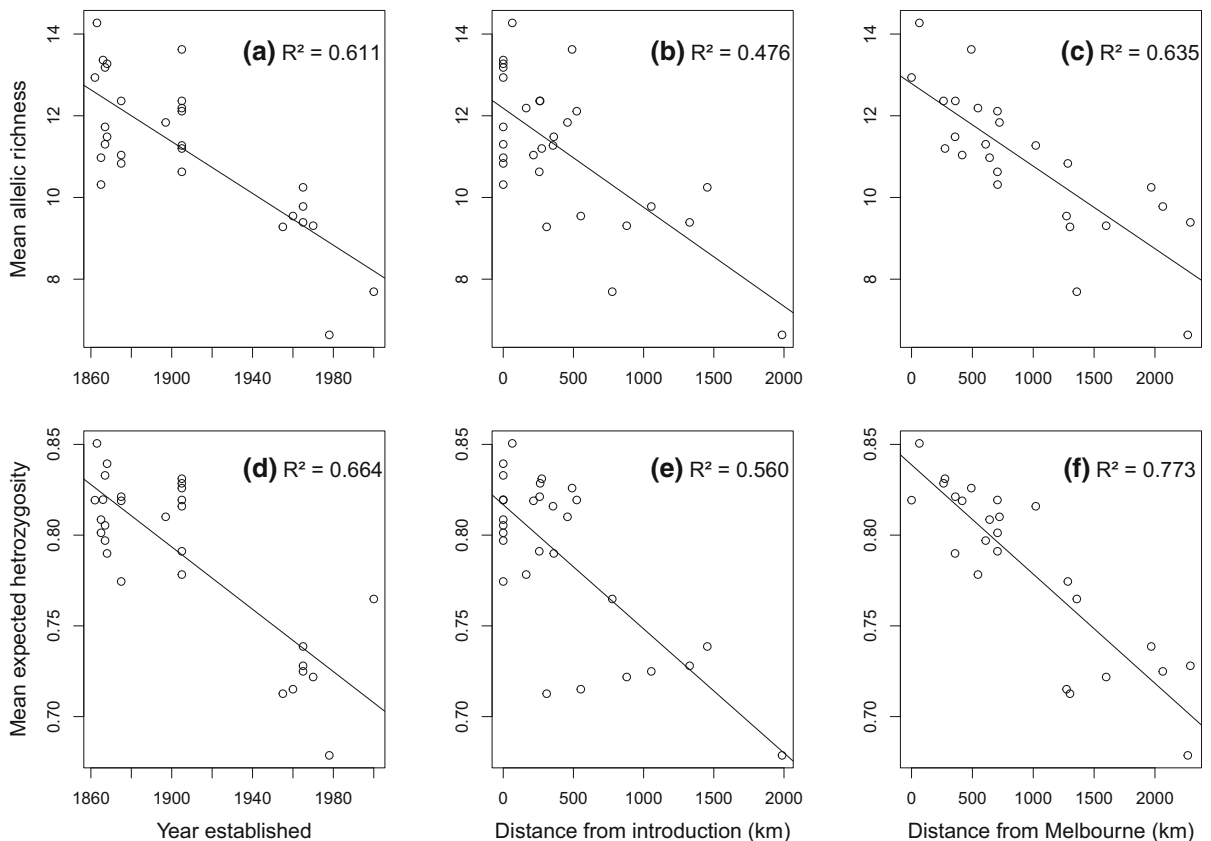
Fig. 4 Genetic diversity has a negative relationship with time and distance from introduction sites. Both allelic richness and expected heterozygosity decline with the recorded year the population was established (**a**, **d**) and the distance from the proposed introduction site (**b**, **e**). However, the same negative relationship exists between these diversity metrics and the distance a population is from the Melbourne introduction site (**c**, **f**). Melbourne was probably the release point for the largest number of sparrows imported to Australia and currently has relatively high levels of genetic diversity. These graphs plot the raw data for each locality, lines of best fit and R^2 values are from linear regressions using the two variables (see Table S5 for details)

2011; Lima et al. 2012; Jensen et al. 2013). However, genetic differentiation was found to be very low (F_{ST} among Finnish populations $= 0.004 \pm 0.001$ s.e.) across a broad distribution of native house sparrow populations in Finland (Kekkonen et al. 2011b). Although house sparrows are generally sedentary, individuals have been shown to naturally disperse up to 50 km in an island archipelago off of Norway (Tufto et al. 2005). The sparrow, can also disperse long distances across uninhabitable landscape to reach new human settlements by hitchhiking on human modes of transport such as trains, trucks and boats (Long 1988). This mode of range expansion has been proposed in introduced sparrow populations in Africa (Schrey et al. 2014) as well as in other invasions that show evidence of long distance unintentional anthropogenic

introductions (Miller et al. 2005; Pascual et al. 2007; Preuss et al. 2015). Long distance dispersal events would have been necessary for the colonization of remote Australian towns. The isolation between rural towns in Australia has led to independent populations and high population differentiation that is characteristic of a meta-population (pairwise F_{ST}, Fig. S1). In contrast, the highly populated areas sampled around Melbourne show lower levels of differentiation (Fig. S1) and the neighbouring populations of Albury and Burrumbuttock show no significant differentiation at this much smaller spatial scale (ca. 30 km apart, $F_{ST} = 0.004$). The pairs of sample sites within the north and south islands of New Zealand also show low differentiation ($F_{ST} = 0.006$ and 0.001 respectively, both n.s.).

We find across all our 25 Australian sampled localities a clear pattern of IBD (Fig. S4a). However, within smaller regions linked to the same population cluster the trend is less strong, possibly because there is more connectivity and gene flow (Fig. S4b and S4d). Another explanation of this result is Simpson's Paradox which describes instances where if data is divided into known categories and analysed separately the original result using the full dataset is no longer supported (Wagner 1982). When we look at IBD across all of Australia the two main categories are comparisons within and between population clusters. There is strong population structure between northern and southern Australia so these pairwise comparisons have high F_{ST} and are also far apart but this genetic differentiation is not necessarily explained by distance alone but also potentially by independent introduction events and low subsequent gene flow between the two populations. Northern Australian sample sites also have lower genetic diversity than southern sites and differences in genetic diversity will affect estimates of genetic differentiation (F_{ST}) with lower diversity increasing F_{ST} (Jakobsson et al. 2013). These differences in genetic diversity can be explained by differences in founder population size as well as genetic drift. There is also the possibility we see genetic diversity decline towards the range edges after a single range expansion from a single introduction point. In Australia, the single point could be Melbourne, which has the highest genetic diversity; we do see a decline in genetic diversity as we move away from the Melbourne area in the South of Australia (Fig. 4). Although the neighbour-joining tree found no significant support for an invasion origin from Melbourne expanding across Australia (Fig. 3). We also see an overall decline in genetic diversity going from the south to the north of Australia (Table S5). In this study, it is most likely that the latitudinal pattern in genetic diversity is best explained by introduction history.

Acclimatisation Societies have documented five main introduction sites in Australia and they were all found to be linked to distinct genetic populations (Fig. 2). These five sites were the capital cities that supported Acclimatisation Societies: Melbourne, Sydney, Hobart, Adelaide and Brisbane. The localities in the Melbourne population show the highest levels of genetic diversity (Table 1 and Fig. S5), which is consistent with the historical records that report the largest numbers of birds being imported into Melbourne (Andrew and Griffith 2016). The Sydney population is reported as being founded by individuals imported in the 1860s from the newly established population in Melbourne. This event would explain why these populations are close to each other in Fig. 2a, but population structure due to founder effects has been maintained possible due to limited gene flow between the two regions. Similarly, if the Hobart population on the island of Tasmania was originally founded by birds sent from Melbourne in 1867 (Andrew and Griffith 2016) then the subsequent genetic drift due to founder effects could have created genetic differentiation with the mainland. The sea barrier between the two islands of New Zealand has also maintained differentiation between the two island populations, that were likely established by separate introductions [see Thomson (1922) and Table S1]. Interestingly the English sample has grouped with the South Island of New Zealand which is also genetically similar to the North Island. The New Zealand populations could be more similar to England than Australian populations because sparrows were imported from England and India to Melbourne with potential successful introgression between sparrows from these two sub-species [*P. d. domesticus* and *P. d. indicus*, (Andrew and Griffith 2016)].

Independent shipments from England were also reported to have successfully introduced sparrows to Brisbane and Adelaide. The sparrows contributing to the primary introduction into Brisbane may include those arriving by ship in 1869 and 12 sparrows sent from Melbourne in 1868 (Andrew and Griffith 2016). There are also clear primary reports that house sparrow populations were established in Brisbane as well as Adelaide in the 1870s, well before sparrows spread naturally out of Victoria without human intervention (Andrew and Griffith 2016). The DAPC (Fig. 2) shows a clear gap between all the sample sites in the Brisbane cluster and the rest of Australia. The Brisbane population also had the most private alleles (three) with a frequency greater than 1%. This is surprising since the Brisbane population also had the lowest number of alleles. These results suggest that the Brisbane population is more than just a subset of the genetic diversity found in the south. However, sequential founder events and genetic drift can also make rare alleles much more common so this population structure could be explained by demographic

bottlenecks during colonization (Excoffier et al. 2009a). The Melbourne and Adelaide population clusters are not as differentiated, indicating more connectivity within the southern half of the species distribution in Australia, if there was a successful independent introduction to Adelaide in the 1860s. It is also possible that both the Adelaide and Melbourne introductions had similar source populations.

The complexity of invasion histories and the demographic bottlenecks experienced by invasive species can provide unique challenges for studies of local adaptation. In parallel with genetic changes due to local adaptation, there are a number of neutral processes driving population subdivision such as sequential colonization events (Peter and Slatkin 2015), small population size/inbreeding (Keller and Waller 2002) and admixture (Orsini et al. 2013). The recent development of genome scan methods and the identification of confounding genetic signals for identifying loci being acted on be selection, adds new value to the study of genetic population structure in invasive species (Excoffier et al. 2009b; Günther and Coop 2013; de Villemereuil et al. 2014; Lotterhos and Whitlock 2014). Information on population structure and demographic history that is gathered through high resolution genetic sampling, can improve the sampling design of genome scan projects and the informative power of results (De Mita et al. 2013; Rellstab et al. 2015; Francois et al. 2016). A more general benefit of describing population structure in invasive populations is to propose general patterns of genetic differentiation in invasive populations. These observations will help us to draw conclusions about the origin and dynamics of biological invasions that have limited historical information.

The large number of reported introduction events of the house sparrows in Australia and New Zealand is probably not unusual for species that were introduced by Acclimatisation Societies to many locations around the world in the mid 1800s (Long 1981; Lever 1992, 2005). The house sparrow has also been used as a model species to study invasion genetics globally (Liebl et al. 2015). Here we have described the population structure of the house sparrow within the last major region of the world that had not been previously subjected to study by genetic markers. The population structure that we have characterised within Australia and New Zealand is consistent with our expectations for this relatively sedentary species that

has gone through reasonably well documented sequential colonization events. Furthermore, the relatively strong structuring that we have characterised suggests that there is reasonable scope for local adaptation to have occurred, even in the relatively short period of time since the introduction, just over 150 years ago. The population of introduced house sparrows around the world remain a good target for further studies of evolution and ecology (Liebl et al. 2015), particularly given the genomic resources that are coming online for this species (Hagen et al. 2013). There is an increasing focus on invasion genetics (Colautti and Lau 2015) and adaptation to urban environments (Johnson and Munshi-South 2017), and our study demonstrates how population structure can be affected by human mediated introductions or by species expanding their range by invading new habitat patches in urban environments. Future studies in this popular area of research may gain insight into evolutionary change by using molecular tools to characterise and account for the invasion history.

References

Anderson TR (2006) Biology of the ubiquitous house sparrow: from genes to populations. University Press, New York

Andrew SC, Griffith SC (2016) Inaccuracies in the history of a well-known introduction: a case study of the Australian House Sparrow (*Passer domesticus*). Avian Res 7:9. https://doi.org/10.1186/s40657-016-0044-3

Baalsrud HT, Saether B-E, Hagen IJ et al (2014) Effects of population characteristics and structure on estimates of effective population size in a house sparrow metapopulation. Mol Ecol 23:2653–2668. https://doi.org/10.1111/mec.12770

Balmford R (1981) Early introductions of birds to Victoria. Vic Nat 98:96–105

Barrett SCH (2014) Foundations of invasion genetics: the Baker and Stebbins legacy. Mol Ecol 24:1927–1941. https://doi.org/10.1111/mec.13014

Bates D, Maechler M, Bolker B, Walker S (2015) lme4: linear mixed-effects models using Eigen and S4. R package version 1.1-8. https://CRAN.R-project.org/package=lme4

Benazzo A, Ghirotto S, Vilaça ST, Hoban S (2015) Using ABC and microsatellite data to detect multiple introductions of invasive species from a single source. Heredity (Edinb) 115:262–272. https://doi.org/10.1038/hdy.2015.38

Blackburn TM (2008) Using aliens to explore how our planet works. Proc Natl Acad Sci 105:9–10

Cavalli-Sforza LL, Edwards AWF (1967) Phylogenetic analysis models and estimation procedures. Am J Hum Genet 19:233–257. https://doi.org/10.2307/2406616

Colautti RI, Lau JA (2015) Contemporary evolution during invasion: evidence for differentiation, natural selection, and local adaptation. Mol Ecol 24:1999–2017

Cole SR, Parkin DT (1986) Adenosine-deaminase polymorphism in house sparrow, *Passer-domesticus*. Anim Genet 17:77–88

Cristescu ME (2015) Genetic reconstructions of invasion history. Mol Ecol 24:2212–2225. https://doi.org/10.1111/mec.13117

Dawson DA, Horsburgh GJ, Krupa AP et al (2012) Microsatellite resources for Passeridae species: a predicted microsatellite map of the house sparrow *Passer domesticus*. Mol Ecol Resour 12:501–523. https://doi.org/10.1111/j.1755-0998.2012.03115.x

De Mita S, Thuillet A-C, Gay L et al (2013) Detecting selection along environmental gradients: analysis of eight methods and their effectiveness for outbreeding and selfing populations. Mol Ecol 22:1383–1399. https://doi.org/10.1111/mec.12182

de Villemereuil P, Frichot É, Bazin É et al (2014) Genome scan methods against more complex models: When and how much should we trust them? Mol Ecol 23:2006–2019. https://doi.org/10.1111/mec.12705

Dlugosch KM, Parker IM (2008) Founding events in species invasions: genetic variation, adaptive evolution, and the role of multiple introductions. Mol Ecol 17:431–449. https://doi.org/10.1111/j.1365-294X.2007.03538.x

Earl DA, VonHoldt BM (2012) STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. Conserv Genet Resour 4:359–361. https://doi.org/10.1007/s12686-011-9548-7

Estoup A, Guillemaud T (2010) Reconstructing routes of invasion using genetic data: Why, how and so what? Mol Ecol 19:4113–4130. https://doi.org/10.1111/j.1365-294X.2010.04773.x

Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software structure: a simulation study. Mol Ecol 14:2611–2620. https://doi.org/10.1111/j.1365-294X.2005.02553.x

Excoffier L, Laval G, Schneider S (2005) Arlequin (version 3.0): an integrated software package for population genetics data analysis. Evol Bioinform Online 1:47

Excoffier L, Foll M, Petit RJ (2009a) Genetic consequences of range expansions. Annu Rev Ecol Evol Syst 40:481–501. https://doi.org/10.1146/annurev.ecolsys.39.110707.173414

Excoffier L, Hofer T, Foll M (2009b) Detecting loci under selection in a hierarchically structured population. Heredity (Edinb) 103:285–298. https://doi.org/10.1038/hdy.2009.74

Francois O, Martins H, Caye K, Schoville S (2016) A tutorial on controlling false discoveries in genome scans for selection. Mol Ecol 25:454–469. https://doi.org/10.1111/mec.13513

Frantz AC, Cellina S, Krier A et al (2009) Using spatial Bayesian methods to determine the genetic structure of a continuously distributed population: Clusters or isolation by distance? J Appl Ecol 46:493–505. https://doi.org/10.1111/j.1365-2664.2008.01606.x

Goudet J (1995) FSTAT (version 1.2): a computer program to calculate F-statistics. J Hered 86:485–486. https://doi.org/10.1093/jhered/esh074

Günther T, Coop G (2013) Robust identification of local adaptation from allele frequencies. Genetics 195:205–220. https://doi.org/10.1534/genetics.113.152462

Hagen IJ, Billing AM, Rønning B et al (2013) The easy road to genome-wide medium density SNP screening in a non-model species: development and application of a 10 K SNP-chip for the house sparrow (*Passer domesticus*). Mol Ecol Resour 13:429–439. https://doi.org/10.1111/1755-0998.12088

Higgins PJ, Peter JM, Cowling SJ (2006) Handbook of Australian, New Zealand and Antarctic birds: Boatbill to starlings, vol 7. Oxford University Press, Melbourne, pp 1102–1126

Jakobsson M, Rosenberg NA (2007) CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. Bioinformatics 23:1801–1806. https://doi.org/10.1093/bioinformatics/btm233

Jakobsson M, Edge MD, Rosenberg NA (2013) The relationship between FST and the frequency of the most frequent allele. Genetics 193:515–528. https://doi.org/10.1534/genetics.112.144758

Jensen H, Moe R, Hagen IJ et al (2013) Genetic variation and structure of house sparrow populations: Is there an island effect? Mol Ecol 22:1792–1805. https://doi.org/10.1111/mec.12226

Johnson MTJ, Munshi-South J (2017) Evolution of life in urban environments. Science 358:eaam8327

Jombart T (2008) Adegenet: a R package for the multivariate analysis of genetic markers. Bioinformatics 24:1403–1405. https://doi.org/10.1093/bioinformatics/btn129

Jombart T, Pontier D, Dufour A-B (2009) Genetic markers in the playground of multivariate analysis. Heredity (Edinb) 102:330–341. https://doi.org/10.1038/hdy.2008.130

Jombart T, Devillard S, Balloux F et al (2010) Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. BMC Genet 11:94. https://doi.org/10.1186/1471-2156-11-94

Kekkonen J, Hanski IK, Jensen H et al (2011a) Increased genetic differentiation in house sparrows after a strong population decline: from panmixia towards structure in a common bird. Biol Conserv 144:2931–2940. https://doi.org/10.1016/j.biocon.2011.08.012

Kekkonen J, Seppä P, Hanski IK et al (2011b) Low genetic differentiation in a sedentary bird: house sparrow population genetics in a contiguous landscape. Heredity (Edinb) 106:183–190. https://doi.org/10.1038/hdy.2010.32

Keller LF, Waller DM (2002) Inbreeding effects in wild populations. Trends Ecol Evol 17:230–241

Kuznetsova A, Brockhoff PB, Christensen RHB (2016) lmerTest: tests in linear mixed effects models. R package version 2.0-32. https://CRAN.R-project.org/package=lmerTest

Langella O (2002) POPULATIONS. Bioinformatics. http://bioinformatics.org/project/?group_id=84

Lee CEE (2002) Evolutionary genetics of invasive species. Trends Ecol Evol 17:386–391. https://doi.org/10.1016/S0169-5347(02)02554-5

Lever C (1985) Naturalized Mammals of the World. Longman, London

Lever C (1992) They dined on eland. Quiller Press Ltd, London

Lever C (2005) Naturalised birds of the world. T & A D Poyser, London, pp 204–218

Liebl AL, Schrey AW, Andrew SC et al (2015) Invasion genetics: lessons from a ubiquitous bird, the house sparrow Passer domesticus. Curr Zool 61:465–476. https://doi.org/10.1093/czoolo/61.3.465

Lima MR, Macedo RHF, Martins TLF et al (2012) Genetic and morphometric divergence of an invasive bird: the introduced house sparrow (Passer domesticus) in Brazil. PLoS ONE. https://doi.org/10.1371/journal.pone.0053332

Long JL (1981) Introduced birds of the world: the worldwide history, distribution and influence of birds introduced to new environments. Universe Books, Campbelltown, pp 372–381

Long JL (1988) Introduced birds and mammals in Western Australia. Agric. Prot. Board West. Aust. Agriculture Protection Board, Perth, pp 13–14

Lotterhos KE, Whitlock MC (2014) Evaluation of demographic history and neutral parameterization on the performance of FST outlier tests. Mol Ecol 23:2178–2192. https://doi.org/10.1111/mec.12725

Marshall TC, Slate J, Kruuk LEB, Pemberton JM (1997) Statistical confidence for likelihood-based paternity. Mol Ecol 7:639–655

Miller N, Estoup A, Toepfer S et al (2005) Multiple transatlantic introductions of the western corn rootworm. Science 310:992. https://doi.org/10.1126/science.1115871

Nakagawa S, Schielzeth H (2013) A general and simple method for obtaining R2 from generalized linear mixed-effects models. Methods Ecol Evol 4:133–142. https://doi.org/10.1111/j.2041-210x.2012.00261.x

Ockendon N, Griffith SC, Burke T (2009) Extrapair paternity in an insular population of house sparrows after the experimental introduction of individuals from the mainland. Behav Ecol 20:305–312. https://doi.org/10.1093/beheco/arp006

Orsini L, Vanoverbeke J, Swillen I et al (2013) Drivers of population genetic differentiation in the wild: isolation by dispersal limitation, isolation by adaptation and isolation by colonization. Mol Ecol 22:5983–5999. https://doi.org/10.1111/mec.12561

Parkin DT, Cole SR (1985) Genetic differentiation and rates of evolution in some introduced populations of the house sparrow, Passer domesticus in Australia and New Zealand. Heredity (Edinb) 54:15–23. https://doi.org/10.1038/hdy.1985.4

Pascual M, Chapuis MP, Mestres F et al (2007) Introduction history of Drosophila subobscura in the New World: a microsatellite-based survey using ABC methods. Mol Ecol 16:3069–3083. https://doi.org/10.1111/j.1365-294X.2007.03336.x

Peakall R, Smouse PE (2006) GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. Mol Ecol Notes 6:288–295. https://doi.org/10.1111/j.1471-8286.2005.01155.x

Peakall R, Smouse PE (2012) GenALEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research-an update. Bioinformatics 28:2537–2539. https://doi.org/10.1093/bioinformatics/bts460

Peter BM, Slatkin M (2015) The effective founder effect in a spatially expanding population. Evolution 69:721–734. https://doi.org/10.1111/evo.12609.This

Preuss S, Berggren Å, Cassel-Lundhagen A (2015) Genetic patterns reveal an old introduction event and dispersal limitations despite rapid distribution expansion. Biol Invasions 17:2851–2862. https://doi.org/10.1007/s10530-015-0915-2

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

R Core Team (2017) R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org

Raymond M, Rousset F (1993) GENEPOP (version 1.2): population genetics software for exact tests and ecumenicism. J Hered 86:248–249

Rellstab C, Gugerli F, Eckert AJ et al (2015) A practical guide to environmental association analysis in landscape genomics. Mol Ecol 24:4348–4370. https://doi.org/10.1111/mec.13322

Rollins LA, Woolnough AP, Wilton AN et al (2009) Invasive species can't cover their tracks: using microsatellites to assist management of starling (Sturnus vulgaris) populations in Western Australia. Mol Ecol 18:1560–1573. https://doi.org/10.1111/j.1365-294X.2009.04132.x

Rosenberg NA (2004) Distruct: a program for the graphical display of population structure. Mol Ecol Notes 4:137–138. https://doi.org/10.1046/j.1471-8286.2003.00566.x

Sætre G-P, Riyahi S, Aliabadian M et al (2012) Single origin of human commensalism in the house sparrow. J Evol Biol 25:788–796. https://doi.org/10.1111/j.1420-9101.2012.02470.x

Sax DF, Stachowicz JJ, Brown JH et al (2007) Ecological and evolutionary insights from species invasions. Trends Ecol Evol 22:465–471. https://doi.org/10.1016/j.tree.2007.06.009

Schrey AW, Grispo M, Awad M et al (2011) Broad-scale latitudinal patterns of genetic diversity among native European and introduced house sparrow (Passer domesticus) populations. Mol Ecol 20:1133–1143. https://doi.org/10.1111/j.1365-294X.2011.05001.x

Schrey AW, Liebl AL, Richards CL, Martin LB (2014) Range expansion of house sparrows (Passer domesticus) in Kenya: evidence of genetic admixture and human-

mediated dispersal. J Hered 105:60–69. https://doi.org/10.1093/jhered/est085

Thomson GM (1922) The Naturalisation of Animals and Plants in New Zealand. University Press, Cambridge

Tufto J, Ringsby T, Dhondt AA et al (2005) A parametric model for estimation of dispersal patterns applied to five passerine spatially structured populations. Am Nat 165:E13–E26

Vangestel C, Mergeay J, Dawson DA et al (2011) Spatial heterogeneity in genetic relatedness among house sparrows along an urban-rural gradient as revealed by individual-based analysis. Mol Ecol 20:4643–4653. https://doi.org/10.1111/j.1365-294X.2011.05316.x

Wagner CH (1982) Simpson's paradox in real life. Am Stat 36:46–48