

## ORIGINAL ARTICLE

# Bottlenecks, population differentiation and apparent selection at microsatellite loci in Australian *Drosophila buzzatii*

JSF Barker<sup>1</sup>, J Frydenberg<sup>2</sup>, J González<sup>3,5</sup>, HI Davies<sup>4</sup>, A Ruiz<sup>3</sup>, JG Sørensen<sup>2</sup> and V Loeschcke<sup>2</sup>

<sup>1</sup>School of Environmental and Rural Science, University of New England, Armidale, New South Wales, Australia; <sup>2</sup>Aarhus Centre for Environmental Stress Research, Ecology and Genetics, Department of Biological Sciences, University of Aarhus, Aarhus C, Denmark;

<sup>3</sup>Departament de Genètica i de Microbiologia, Universitat Autònoma de Barcelona, Bellaterra (Barcelona), Spain and <sup>4</sup>School of Mathematics, Statistics and Computer Science, University of New England, Armidale, New South Wales, Australia

Species colonizing new areas disjunct from their original habitat may be subject to novel selection pressures, and exhibit adaptive genetic changes. However, if colonization occurs through a small number of founders, the genetic composition of the colonized population may differ from that of the original population simply due to genetic drift. Disentangling the effects of founder drift and selection after colonization is crucial to understanding the adaptive process. *Drosophila buzzatii* colonized Australia some 600–700 generations ago, and spread rapidly over a wide geographical range. Genetic variation for 15 microsatellite loci in each of nine populations in eastern Australia was used to estimate the size of the bottleneck, and to determine if any of these microsatellites marked genomic regions subject to recent selection. We

estimate that on its introduction to Australia, *D. buzzatii* went through a moderate bottleneck (approximately 30–40 founders). Linkage disequilibrium was common, both intra-chromosomal and between loci on different chromosomes. Of the 15 loci, 2 showed evidence of selection, one exhibiting local adaptation in different populations and the other balancing selection. We conclude that linkage disequilibria may be far more common in natural populations than is generally assumed, and the loci apparently affected by selection may well be marking selection in large genome regions including many loci that are not necessarily closely linked.

*Heredity* (2009) **102**, 389–401; doi:10.1038/hdy.2008.127; published online 14 January 2009

**Keywords:** genotype–climate associations; bottlenecks; colonization; linkage disequilibrium; latitudinal transect; cactophilic *Drosophila*

## Introduction

Species colonizing new areas disjunct from their original habitat may be subject to novel selection pressures, and exhibit adaptive genetic changes. However, if colonization occurs through a small number of founders, the genetic composition of the colonized population may differ from that of the original population simply due to genetic drift. Any selection will affect particular regions of the genome, whereas demography, migration and the mating system will affect the whole genome in the same way (Cavalli-Sforza, 1966). Although genetic drift will affect all loci in the genome, the consequences for different loci may be different because of stochastic effects. Thus, the problem still is to disentangle the effects of drift due to a small founding population from those of selection after colonization. A neutral microsatellite may show apparent selective effects if it is in linkage disequilibrium (LD) with

other loci subject to selection. When populations are exposed to different environments, diversifying selection at some loci can lead to increased levels of differentiation for those loci relative to neutral expectations, whereas spatially uniform or balancing selection affecting particular loci can lead to decreased differentiation relative to neutral expectations.

In recent years, loci affected by selection have been identified using multilocus simulation tests. These methods, which use summary statistics to identify ‘outlier’ loci (those affected by selection or selective sweeps), may be divided into two classes of interlocus comparisons: (1) those based on levels of differentiation between populations, and (2) those based on comparisons of diversity within populations (Storz, 2005). In addition, multiple regression methods may be used to test for associations between allele frequencies (and heterozygosities) and spatial and climatic variables. Although associations with spatial variables could be due to drift and/or migration, any associations with climatic variables after correction for geographic location suggest effects of differential selection in different populations (Endler, 1986).

Here we apply these methods to investigate possible adaptive differences among *Drosophila buzzatii* popula-

Correspondence: Professor JSF Barker, 114 Cooke Road, Witta, Maleny, Queensland 4552, Australia.

E-mail: sbarker@une.edu.au

<sup>5</sup>Current address: Department of Biology, Stanford University, Stanford, CA 94305, USA.

Received 26 August 2008; revised 22 November 2008; accepted 7 December 2008; published online 14 January 2009

tions in Australia. *D. buzzatii* is a cactophilic species, feeding and breeding in the rotting tissues of *Opuntia* cactus (prickly pear) species, and is restricted to this niche in Australia (Barker and Mulley, 1976). The Chaco region of Argentina is the presumed center of origin of *D. buzzatii* (Vilela *et al.*, 1980), and its introduction to Australia was apparently from this region in the period 1931–1936 (Barker, 1982; Barker *et al.*, 1985), during the program for the biological control of prickly pear (Mann, 1970). *Cactoblastis cactorum* (Berg) (Lepidoptera: Pyralidae), which was introduced from Argentina in 1925 as a potential control agent, proved extremely effective—the larvae destroying plant tissue, leading to rot and death of the plants. By the time of the inadvertent introduction of *D. buzzatii*, *C. cactorum* had been released extensively throughout the *Opuntia* infestation (Dodd, 1940). By 1931, the density of *C. cactorum* was estimated at  $2.5 \times 10^7$  larvae per hectare, over the thousands of square kilometers of dense prickly pear (Dodd, 1940). Thus, *D. buzzatii* then had extremely large areas of suitable habitat, probably spread rapidly to all the *Opuntia*-infested area, and must have had an enormous population size (Barker, 1982; Sokal *et al.*, 1987). By 1940, as the *Opuntia* was controlled and its distribution reduced more or less to the habitat islands found today, *D. buzzatii* also contracted to the spatially isolated populations that still exist. Colonization of this extensive territory in Australia is likely to have been accompanied by genetic adaptation to the new environments.

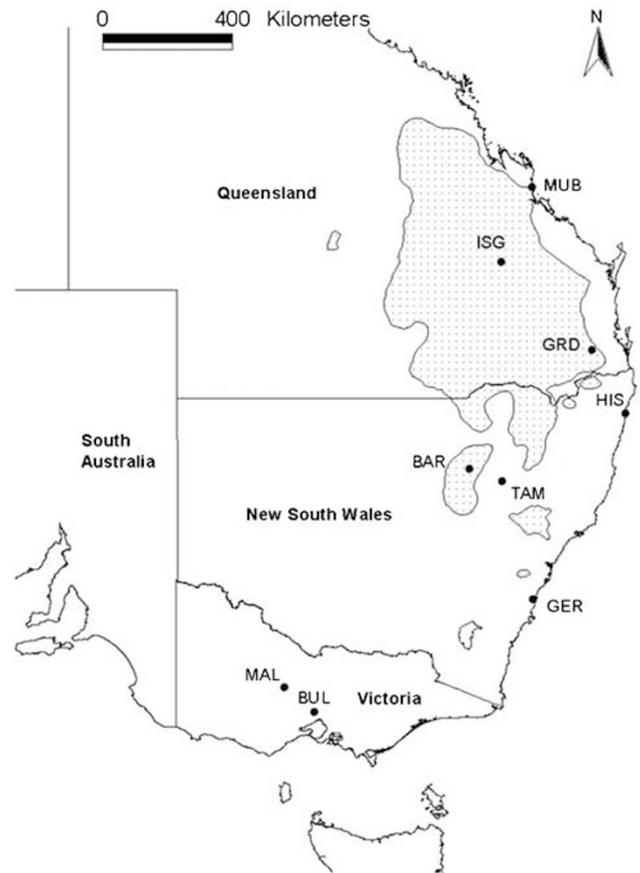
Our objectives, using 15 microsatellite loci, were to assess genetic variation and population structure, to determine the size of the Australian founding population, and to determine if any of the microsatellite loci mark genome regions showing selective effects. As such effects may depend on loci located within inversions, which are known to vary in frequency among Australian populations (Knibb and Barker, 1988), we mapped the microsatellites used in this study.

## Materials and methods

Flies were collected at nine localities in eastern Australia (Table 1; Figure 1) in March–April 2002, by netting from banana baits, and stored in absolute alcohol.

## DNA extraction and microsatellite analysis

DNA extraction and genotyping methods were as given by Frydenberg *et al.* (2002). We used the 10 microsatellites identified by Frydenberg *et al.* (2002) (annealing temperature in parentheses): Db052 (50 °C), Db087 (45 °C), Db122 (55 °C), Db142 (55 °C), Db223 (60 °C), Db225 (58 °C), Db290 (45 °C), Db411 (60 °C), Db493 (60 °C) and



**Figure 1** Map of eastern Australia showing collection localities, and the distribution of the main *Opuntia* infestation in 1920 (hatched area).

**Table 1** Geographical coordinates for the Australian populations (listed from north to south) and genetic variability of each population (s.e. in parentheses)

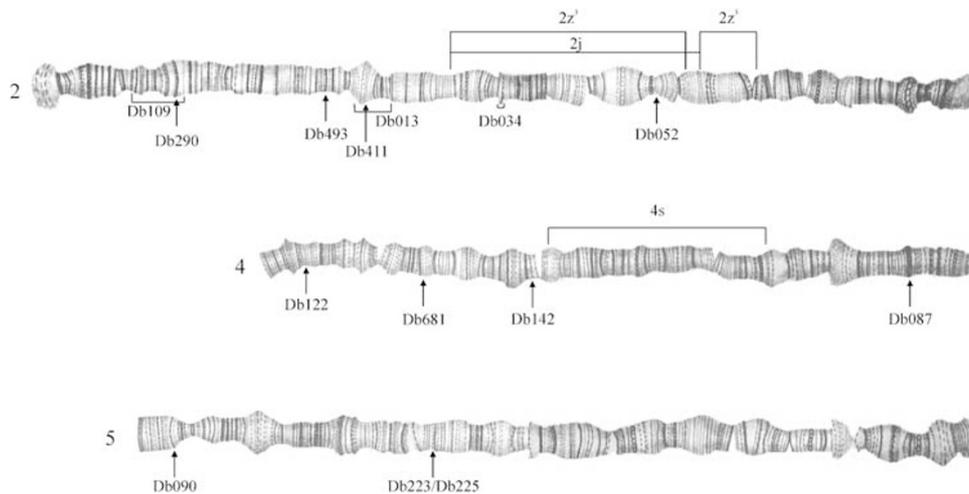
Population (code)	Latitude	Longitude	Sample size <sup>a</sup>	Mean number of alleles/locus	Mean heterozygosity	
					Direct count	Hardy–Weinberg expected
<i>Australia</i>						
Mulambin Beach (MUB)	23.188 S	150.789 E	30	8.13 (0.86)	0.637 (0.048)	0.650 (0.046)
Isla Gorge (ISG)	25.247 S	149.941 E	30	8.00 (0.93)	0.697 (0.032)	0.708 (0.032)
Grandchester (GRD)	27.678 S	152.445 E	30	7.67 (0.74)	0.677 (0.033)	0.685 (0.030)
Hickey Island (HIS)	29.434 S	153.357 E	30	6.53 (0.65)	0.656 (0.040)	0.659 (0.030)
Baradine (BAR)	30.945 S	149.067 E	30	6.87 (0.70)	0.653 (0.047)	0.652 (0.033)
Tambar Springs (TAM)	31.307 S	149.943 E	30	7.87 (0.94)	0.633 (0.040)	0.631 (0.040)
Gerongar Point (GER)	34.555 S	150.811 E	30	5.33 (0.56)	0.602 (0.046)	0.562 (0.043)
Maldon (MAL)	36.983 S	143.950 E	30	5.07 (0.47)	0.585 (0.044)	0.590 (0.038)
Bulla (BUL)	37.654 S	144.769 E	30	5.27 (0.50)	0.498 (0.051)	0.502 (0.049)
<i>Argentina</i>						
Otamendi (OTA)	34.233 S	58.867 W	20	8.47 (0.99)	0.685 (0.053)	0.743 (0.051)
Catamarca (CAT)	28.600 S	65.800 W	20	10.80 (1.10)	0.770 (0.044)	0.794 (0.041)

<sup>a</sup>Number of individuals sampled.

**Table 2** Properties of the five new microsatellites

Name	Repeat motif	Primer sequence	Product size
Db003	(CA) <sub>3</sub> +2bp+(CA) <sub>2</sub> +12bp+(CA) <sub>10</sub>	F-CGGCTCCCAATTGCATACT R-TGCCAAATCTGCGCAAGTAT	256
Db013	(CA) <sub>16</sub>	F-TACATGTTGCACAACATTGA R-GATAGCAATGCCAGCAATTC	92
Db034	(CACAA) <sub>2</sub> (CA) <sub>9</sub>	F-TCAGGTAGCCTAATTGGCTG R-AGCAAAGCGTGTGGTTAACG	120
Db090	(CA) <sub>11</sub> +2bp+(CA) <sub>3</sub>	F-TACGCGAATGCTAAGCCAAG R-CGTATACGCAGCGTGTGGTT	195
Db109	(CT) <sub>8</sub> +8bp+(CT) <sub>2</sub> (CA) <sub>2</sub> (CT) <sub>9</sub>	F-GGCAACGATTGCGCATACGC R-ACTTGTACATGTGCAGCCAC	149

Product size is the size (bp) of the sequenced clone.



**Figure 2** Localization of the microsatellite clones (except Db003 in chromosome 5) in the *D. buzzatii* polytene chromosomes, with inversion positions also marked. Db003 is located between the genes *Obp56g* and *Toll-7* in scaffold 6496, which corresponds to *D. buzzatii* chromosome 5. Further localization is not possible, as few markers are available. Microsatellites Db013, Db034 and Db109 are located in scaffold 6540, which corresponds to *D. buzzatii* chromosome 2. Db013 is in the chromosomal segment delimited by bands A2f-C4h between genes *jar* and *Atpalpha*. Db034 is located in the chromosomal segment delimited by bands F1h-F2a between genes *amon* and *fkf*. Finally, Db109 is located in the chromosomal segment delimited by bands B2c-B4d between genes *ry* and *Act87E*.

Db681 (40 °C), together with 5 additional ones (Table 2): Db003 (60 °C), Db013 (60 °C), Db034 (60 °C), Db090 (52 °C) and Db109 (60 °C). For some analyses, we included data for two populations in Argentina (Otamendi and Catamarca), using the results for the first 10 microsatellites from Frydenberg *et al.* (2002), together with the 5 additional ones genotyped for this study.

#### *In situ* localization of the microsatellite clones

Most of the microsatellite clones were *in situ* localized to the polytene chromosomes of *D. buzzatii*. DNA from clones Db052, Db090, Db122, Db142, Db223, Db225, Db290, Db411 and Db493 was labeled using the PCR DIG labeling mix (Roche, Basel, Switzerland) and M13 universal forward and reverse primers. For microsatellite clones Db087 and Db681, similarity searches against the *D. mojavensis* genome sequence database (<http://rana.lbl.gov/drosophila/> as of 6 December 2004) were carried out using their sequences as a query. The sequence with the highest similarity for each of the two microsatellites was used to design primers and to amplify fragments of ~1.5 kb (5'-GCTACATGTGGTACCATAAC-3'; 5'-ATAA

CATCATGTGGGGTCTG-3') and ~2 kb (5'-GTGGAACA TGGAAATTGCTAG-3'; 5'-CACGTGGACATCAAGACTG A-3') including the microsatellite sequences of clones Db681 and Db087, respectively. DNA was amplified using genomic DNA of *D. mojavensis* (as a control) and *D. buzzatii*, and the latter PCR product was used as a probe. Fluorescent *in situ* hybridization was carried out as in González *et al.* (2000), with detection using anti-digoxigenin antibodies labeled with fluorescein-5-isothiocyanate. Chromosomal localization of the hybridization signals (Figure 2) was determined using the cytological map of *D. buzzatii* (González *et al.*, 2005). Microsatellites Db003, Db013, Db034 and Db109 were blasted against the *D. mojavensis* genome sequence, and localized using data on gene localization in *D. buzzatii* (Ranz *et al.*, 2003).

#### Allelic frequency, heterozygosity and linkage disequilibrium

Genotype and allele frequencies were estimated using GENEPOP version 3.4 (Raymond and Rousset, 1995; GENEPOP data file for all 11 populations available as Supplementary Table S1), and alleles per locus and

observed and expected heterozygosity (gene diversity) were estimated using GENECLASS2 (Piry *et al.*, 2004). Tests for deviations from Hardy–Weinberg equilibrium were carried out using the exact tests of GENEPOP (default values for the Markov chain method). Significance levels for each test were determined by applying to the probability estimates calculated by GENEPOP, the sequential Bonferroni procedure (Hochberg, 1988; Lessios, 1992) over loci within each population. The number of alleles per locus, and observed and expected heterozygosity were compared among populations with the Kruskal–Wallis nonparametric test (Sokal and Rohlf, 1981). Pair-wise differences among populations for expected heterozygosity were tested using a *t*-test on arcsine transformed values (Archie, 1985), with the sequential Bonferroni procedure applied over the set of eight comparisons for each population.

Pair-wise linkage disequilibria ( $D'$ ) between loci within each population were estimated using PowerMarker version 3.25 (Liu and Muse, 2005). Statistical significance was evaluated using the exact tests implemented in PowerMarker, with *P*-values obtained using both permutation and the Markov chain Monte Carlo approaches. Both approaches gave similar results, and only the former are presented. *P*-values were not adjusted for multiple comparisons. In addition to mean  $D'$  estimates, we use for each population the percentage of locus pairs that had significant ( $P < 0.05$ )  $D'$  values. As 5% of pair-wise LD are expected by chance to be significant, higher percentages indicate more LD than would be expected (Schug *et al.*, 2007).

#### Population differentiation

As one locus in one population showed a significant deviation from Hardy–Weinberg equilibrium, both genotypic and genic differentiation among populations were tested using GENEPOP, for overall and pair-wise differentiation (default values for the Markov chain method). The sequential Bonferroni procedure was applied over population pairs for the latter in determining significance levels. *F*-statistics (Weir and Cockerham, 1984) and their significance were determined using FSTAT version 2.9.3 (Goudet, 2001), not assuming Hardy–Weinberg equilibrium and with 5000 iterations, and the sequential Bonferroni procedure was applied over loci to determine significance levels. Isolation by distance was tested using estimates of  $F_{ST}$  (Weir and Cockerham, 1984) for each pair of populations, with pair-wise  $F_{ST}/(1-F_{ST})$  values regressed on log (geographic distance) between each pair of populations (Rousset, 1997), and the significance of the association was determined using Mantel's (1967) permutation test.

#### Population history: bottlenecks

On its introduction to Australia, *D. buzzatii* is presumed to have been subject to a bottleneck (Barker *et al.*, 1985; Halliburton and Barker, 1993), followed by a rapid and massive expansion in population size. Subsequently, populations were reduced to a patchy distribution, but with substantial variation in the number of plants (and rotting tissue) at each locality. The number of founders of the colonized Australian population was estimated using the maximum likelihood method of Ramirez *et al.* (2006).

This method computes the likelihood (*L*) that a particular combination of alleles is observed in the colonized population as a function of the number of founder gametes ( $2N$ ), given the allelic frequencies in the source population. Then  $2N$  is estimated as the value that maximizes the likelihood function (*L*). The variance of the estimate is calculated as the inverse of the amount of information, which is equal to the second derivative of  $\ln L$ .

The BOTTLENECK (Cornuet and Luikart, 1996) and M (Garza and Williamson, 2001) programs were used to determine if any signal of past bottlenecks could be detected. The BOTTLENECK analysis tests whether there is excess heterozygosity as compared with the heterozygosity expected from the observed number of alleles at each locus, assuming mutation–drift equilibrium. Bottlenecked populations are predicted to show an excess of heterozygosity, as the number of alleles is more severely affected than heterozygosity by a bottleneck in population size. Expected values were determined using the two-phase mutation model (Piry *et al.*, 1999), with model options 80% single-step mutations, a variance among multiple steps of 12, and 5000 iterations. The probability of significant heterozygosity excess was determined using Wilcoxon's signed-rank test.

*M* is defined as the ratio of the total number of alleles (*k*) to the overall range in allele size (*r*). With genetic drift in populations reduced in size, the loss of any allele will reduce *k*, but only the loss of the largest or smallest allele will reduce *r*. Thus, *M* is expected to be smaller in recently reduced populations than in equilibrium populations. The ratio *M* is estimated for each locus, averaged over loci, and then statistically tested by comparing the estimate to a critical value ( $M_c$ ), estimated for a specific mutation model using 10 000 replicates. The parameters of the model used here were fraction of mutations larger than single step = 0.2, mean size of non-single-step mutations = 3.5 and  $\theta = 4N_e\mu$  (where  $N_e$  = effective population size at equilibrium before the bottleneck and  $\mu$  = mutation rate) equal to 2 (assuming  $N_e = 5000$ ,  $\mu = 10^{-4}$  or  $N_e = 50 000$ ,  $\mu = 10^{-5}$ ). The value assumed for  $\theta$  is consistent with the mean expected heterozygosity ( $H_e$ ) of the Argentine populations. Although Piry *et al.* (1999) recommend 95% single-step mutations, and Garza and Williamson (2001) recommend 90%, we used 80% as 10 of the 15 microsatellites are interrupted or compound. Although there is uncertainty about the appropriate percentage of single-step mutations, our test is conservative as higher percentages and also smaller values of  $\theta$  (lower  $N_e$  of the pre-bottleneck Argentine population) increase the value of  $M_c$  (Garza and Williamson, 2001).

#### Geographical variation in genetic diversity and allele frequency: Australian populations

For each locus, effects of geographical location and climate on expected heterozygosity and allele frequencies were analyzed. In Australia, *D. buzzatii* has been collected at 97 localities, and for each of these, climatic variables were estimated using the BIOCLIM program of the ANUCLIM 5.1 package (Houlder *et al.*, 2000). With the position of a locality described by latitude, longitude and elevation, all 35 climatic variables that can be

produced by BIOCLIM were estimated for each locality (Barker *et al.*, 2005). Principal component analysis applied to these data for each locality (SAS Institute, 1985) provided a summary of the climatic environment for each. The first four principal components accounted for 93% of the variation, and the scores for the nine localities sampled here were considered for use. But with only nine localities, degrees of freedom for multiple regression analyses were limiting. Following preliminary testing, the model used included latitude, longitude and the first two principal components (which accounted for 75% of the climatic variation), and analyses were carried out using the statistical package R (R Development Core Team, 2003). Again with only nine localities, finding a suitable model is problematical. Box-Cox plots were computed for each dependent variable to check possible transformations, but none was found to be appropriate. Normal QQ plots of residuals were used to check the assumption of normality for the fitted models. For the analysis of allele frequencies at each locus, only the two alleles that were at highest frequency over all localities were used. Thus, a significant effect for only one of these alleles means a reciprocal effect for the pooled remaining alleles. All tests for individual terms were adjusted for other terms in the model (type II tests).

#### Multilocus simulation tests for selection

Two multilocus selection tests were used, referred to as the Schlötterer and Beaumont tests. The Schlötterer tests (Schlötterer, 2002; Kauer *et al.*, 2003) use test statistics based on the ratio of observed variances in repeat number (of microsatellites) (ln RV test) or expected heterozygosity (ln RH test) in two groups of populations. The rationale for the tests is that positive selection at a locus will lead to a reduction in variability at the selected locus and flanking regions, so that a microsatellite locus linked to a selected locus is expected to have reduced variability compared to neutral expectations (Schlötterer, 2002). We compared the colonized populations in Australia with two populations in Argentina, using the expected heterozygosity (ln RH test) only, as Kauer *et al.* (2003) show the ln RH test to have higher power to detect selected loci than the ratio of observed variances in repeat number, due to smaller variance of the former, and sensitivity of the latter to non-stepwise mutations. Further, heterozygosity is expected to return to its expected equilibrium more rapidly than variance in allele size (Kimmel *et al.*, 1998), so that ln RH should have more power to detect selective sweeps that have occurred in the relatively recent past, such as during the colonization of Australia. The Beaumont test aims to identify outlier loci (low or high levels of genetic differentiation) by comparing observed  $F_{ST}$  to a null distribution, conditional on heterozygosity, generated by the coalescent simulation model of Beaumont and Nichols (1996). We used a 100 island model of population structure and assumed the stepwise mutation model, with 50 000 paired values of  $F_{ST}$  and heterozygosity generated using the program FDIST2 (<http://www.rubic.rdg.ac.uk/~mab/>). Results are visualized by plotting  $F_{ST}$  vs heterozygosity for each locus, together with the 0.025, 0.50 and 0.975 quantiles of the null distribution of  $F_{ST}$ .

## Results

### Gene diversity and linkage disequilibrium

All loci were polymorphic in all populations (allele frequencies are available in Supplementary Table S2). Some populations showed markedly different allele frequencies for some loci, with the overall most frequent allele not the most frequent in one or more populations for 7 of the 15 loci, and 31 of the alleles were unique to one population. Of these, 11 were unique to Mulambin Beach, with 4 unique of the 17 alleles at Db290.

The mean numbers of alleles per locus (Table 1) were significantly different among populations ( $P < 0.01$ ), but not significantly different among the three southern populations, nor among the six northern populations. The three southern populations generally had significantly fewer alleles. Observed heterozygosities for each population were not significantly different, but expected heterozygosity (gene diversity) differed significantly among populations ( $P < 0.05$ ). Tests for pair-wise differences in expected heterozygosity were not significant among the three southern populations, nor among the six northern populations. After Bonferroni correction, pair-wise differences were significant for Isla Gorge and Grandchester with Bulla. Only Db109 in Baradine showed significant ( $P < 0.05$ ) deviation from Hardy–Weinberg equilibrium ( $F = 0.136$ ) after Bonferroni correction.

Linkage disequilibrium is present in the Australian populations (Table 3), with mean  $D' = 0.260$ , but significantly higher ( $P < 0.05$ ) in the northern populations

**Table 3** LD analyses

(a) Mean LD for all pair-wise comparisons in each population, and percentage of pair-wise comparisons with probability  $< 5\%$

Population	Mean $D'$	% $D'$ significant ( $P < 0.05$ )
MUB	0.482	15.2
ISG	0.469	14.3
GRD	0.451	18.1
HIS	0.430	16.2
BAR	0.479	28.6
TAM	0.441	9.5
GER	0.415	20.0
MAL	0.388	21.9
BUL	0.405	14.3
OTA	0.636	22.9
CAT	0.671	8.6

(b) Locus pairs significant in four or more Australian populations

Locus pairs (no. of populations)	Chromosome(s)
Db034/Db052 (7)	2
Db034/Db290 (6)	2
Db052/Db090 (6)	2 and 5
Db052/Db493 (6)	2
Db034/Db122 (5)	2 and 4
Db052/Db411 (5)	2
Db052/Db122 (4)	2 and 4
Db052/Db681 (4)	2 and 4
Db109/Db681 (4)	2 and 4
Db290/Db681 (4)	2 and 4

(0.274) than in the southern ones (0.232). Five populations had more than 5% significant pair-wise comparisons, whereas 58 (55.2%) of the 105 pair-wise  $D'$  were not significant in any population. A maximum of three populations had significant LD for any one locus pair and 36 locus pairs were significant in only one population. Considering all pair-wise tests ( $105 \times 9$  populations = 945), 59 (6.24%) were significant, of which 31 were between loci on the same chromosome, 21 between loci on different chromosomes and 7 involving the unlocated Db003 (Table 3). Of the 23 significant tests for loci on chromosome 2, 14 involved the two loci (Db034 and Db052) that are located within the inversion segment. Similarly, for loci on different chromosomes, three of the eleven for chromosomes 2 and 4, and two of the six for chromosomes 2 and 5 involve these two loci. Mean  $D'$  is significantly higher ( $P < 0.05$ ) for the Argentine populations (0.462), but neither population had more than 5% significant  $D'$  values. As in the Australian populations, a high proportion (7 out of 8, 87.5%) of the significant LD involve loci on different chromosomes (Table 3).

#### Population differentiation

Analysis of F-statistics (Table 4) showed highly significant population differentiation ( $F_{ST}$ ) for all loci and overall, whereas no  $F_{IS}$  estimates were significantly different from zero. Genic and genotypic differentiation among all populations also were highly significant ( $P < 0.001$ ) for each locus and overall.

All pair-wise  $F_{ST}$  estimates were significant, except for TAM/ISG. Mean pair-wise  $F_{ST}$  estimates were lower for the northern populations ( $0.016 \pm 0.010$ ) than for the southern ( $0.094 \pm 0.023$ ), and means for the northern vs the Argentinian populations ( $0.130 \pm 0.012$ ) also were lower than for the southern vs Argentine ( $0.211 \pm 0.030$ ). The regression of  $F_{ST}/(1-F_{ST})$  on  $\log$  (geographic distance) for the Australian populations ( $b = 0.019$ ) was

**Table 4** F-statistics analysis, with significance determined by permutation tests in the FSTAT program

Locus <sup>a</sup>	$F_{IT}$ (F)	$F_{ST}$ ( $\theta$ )	$F_{IS}$ (f)
Db003	0.115 (0.039)**	0.060 (0.034)***	0.059 (0.019)
Db013	0.083 (0.049)**	0.122 (0.034)***	-0.044 (0.027)
Db034	0.177 (0.058)**	0.117 (0.034)***	0.067 (0.044)
Db052	0.086 (0.050)*	0.087 (0.027)***	-0.001 (0.038)
Db087	0.034 (0.033)	0.039 (0.016)***	-0.006 (0.026)
Db090	0.091 (0.055)**	0.067 (0.030)***	0.025 (0.038)
Db109	0.088 (0.051)**	0.092 (0.030)***	-0.005 (0.039)
Db122	0.015 (0.052)	0.085 (0.031)***	-0.077 (0.047)
Db142	0.108 (0.038)**	0.092 (0.035)***	0.017 (0.024)
Db223	0.186 (0.060)**	0.135 (0.052)***	0.059 (0.038)
Db225	0.123 (0.037)**	0.101 (0.035)***	0.025 (0.029)
Db290	0.148 (0.063)**	0.161 (0.045)***	-0.015 (0.040)
Db411	0.130 (0.029)**	0.082 (0.028)***	0.052 (0.028)
Db493	0.186 (0.055)**	0.127 (0.051)***	0.068 (0.030)
Db681	0.071 (0.047)*	0.087 (0.027)***	-0.017 (0.035)
Mean <sup>b</sup>	0.112 (0.013)***	0.097 (0.009)***	0.016 (0.010)*

These statistics are defined (Weir and Cockerham, 1984) as the correlations between pairs of alleles: (1) within individuals (F), (2) between individuals in the same population ( $\theta$ ) and (3) within individuals within one population (f), and are analogous to Wright's (1951)  $F_{IT}$ ,  $F_{ST}$  and  $F_{IS}$ .

\* $P < 0.05$ , \*\* $P < 0.01$ , \*\*\* $P < 0.001$ .

<sup>a</sup>S.d. in parentheses, estimate from jackknife over populations.

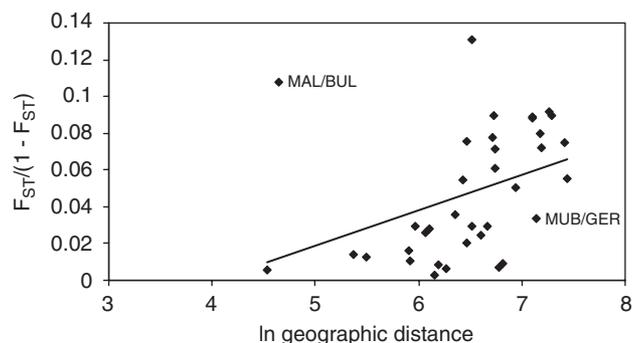
<sup>b</sup>S.d. in parentheses, estimate from jackknife over loci.

significant ( $P < 0.01$ ), indicating isolation by distance, but some populations that are widely separated were not different (for example, MUB/GER) whereas others that are geographically closer were very different (for example, MAL/BUL) (Figure 3).

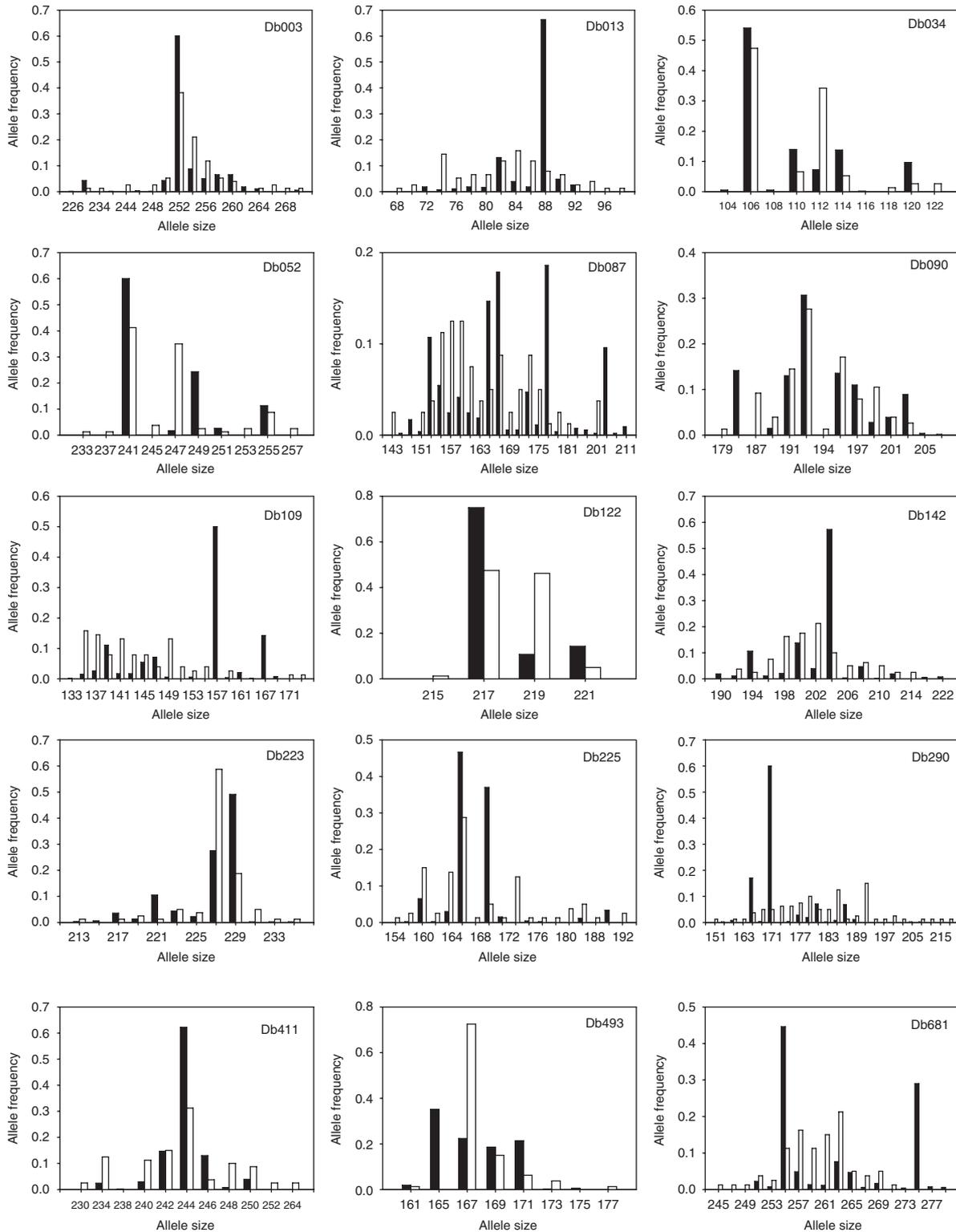
#### Estimation of the number of founders and bottleneck tests

The allele frequency distributions of the overall Australian and Argentinian samples (Figure 4) show the reduced variability of the former as compared with the latter, consistent with the postulated founder effect at colonization. Here we estimate the number of founders using the method of Ramirez *et al.* (2006). The Australian sample (all nine populations, 540 genomes) included 174 alleles, whereas the source sample (two Argentinian populations, 80 genomes) included 187 alleles. The number of alleles shared between the two samples was 134, with a further 40 exclusive to Australia and 53 exclusive to Argentina. Assuming the Argentinian sample as representative of the source population, the estimated number of founder gametes ( $2N \pm s.d.$ ) is  $32.4 \pm 3.6$ . But this neglects the 40 alleles exclusive to Australia—implicitly assuming that they were produced by mutation after colonization. Alternatively, we consider that these 40 alleles are present in Argentina, but not included in our sample of only 80 genomes. This is more appropriate, given that the sample size is much larger for Australia than for Argentina (540 vs 80). Thus, taking the combined Australia + Argentina sample of 227 alleles as a better representation of the source population, the estimated number of founder gametes is  $74.8 \pm 8.3$ . Given the small size of the Argentinian sample, alleles additional to those detected are likely to exist there, so that the number of founder gametes would be less than our estimate. Thus, this analysis indicates a moderate bottleneck of about 30–40 individual founders on colonization of Australia.

The three southern Australian populations showed lower genetic variability than the six northern populations (Table 1), and of the 174 alleles in the total Australian sample, 168 were present in the northern populations and 112 in the southern. The number of alleles shared between the two was 106, with 62 exclusive to the north and 6 to the south. This suggests a secondary colonization of the south from the north, supporting the expectation from historical records that show that *C. cactorum* (and thus habitat for *D. buzzatii*) was deliberately spread only through central and



**Figure 3** Isolation by distance—relationship of population differentiation ( $F_{ST}$ ) to geographic distance for the nine Australian populations.



**Figure 4** Allele frequency distributions for the combined Australian populations (black), and for the combined Argentinian populations (white). Allele size = base pairs. Some rare alleles are omitted for Db142 and Db290 to clarify presentation.

northern New South Wales and Queensland (Dodd, 1940). Assuming the northern population as the source, the estimated number of founder gametes is  $43.4 \pm 6.0$ . But again this assumes that the six alleles exclusive to the southern populations arose after colonization. Thus, taking the total Australian sample of 174 alleles as

representative of the source population, the estimated number of founder gametes of the southern population is  $61.7 \pm 8.6$ , indicating a secondary bottleneck size of about 25–35 founding individuals.

BOTTLENECK analysis (Table 5) showed heterozygosity deficiency for all of the Australian populations.

**Table 5** Tests for past bottlenecks in population size using the BOTTLENECK and M programs: probability values for tests in the BOTTLENECK analysis, and average values of the M ratio for each population and critical values ( $M_c$ ).

Population	Bottleneck		M ratio	
	% loci heterozygosity deficiency	Two-tailed Wilcoxon P-value	M	$M_c$
MUB	0.667	0.008	0.652*	0.675 <sup>a</sup>
ISG	0.667	0.208	0.692	
GRD	0.800	0.015	0.704	
HIS	0.667	0.073	0.664*	
BAR	0.667	0.135	0.647*	
TAM	0.800	0.002	0.659*	
GER	0.600	0.095	0.559***	
MAL	0.667	0.389	0.612**	
BUL	0.600	0.055	0.589**	
OTA	0.400 <sup>b</sup>	0.978	0.693	0.667 <sup>c</sup>
CAT	0.733	0.073	0.733	

\* $P < 0.05$ , \*\* $P < 0.01$ , \*\*\* $P < 0.001$ .<sup>a</sup>Critical value for Australian populations, sample size = 60.<sup>b</sup>Observed heterozygosity excess; observed heterozygosity deficiency for all other populations.<sup>c</sup>Critical value for Argentine populations, sample size = 40.**Table 6** Multiple regression analyses that were significant for effects of geographical location (latitude and longitude) and climatic variables (PC1 and PC2) on expected heterozygosity and allele frequencies at each locus

Locus/allele	Regression coefficients (s.e.)			
	Latitude	Longitude	PC1	PC2
<i>Expected heterozygosity</i>				
Db090	0.044 (0.007)**	-0.032 (0.006)**	0.051 (0.008)**	0.077 (0.011)**
Db142	-0.148 (0.017)**	0.144 (0.015)**	-0.172 (0.019)**	-0.220 (0.026)**
Db681	-0.069 (0.012)**	0.062 (0.011)**	-0.079 (0.014)**	-0.098 (0.019)**
<i>Allele frequencies<sup>a</sup></i>				
Db090/193 (0.282) <sup>b</sup> /195 (0.195)	-0.108 (0.024)*	0.092 (0.022)*	-0.125 (0.028)*	-0.180 (0.037)**
Db142/204 (0.572) /200 (0.137)	0.139 (0.013)**	-0.130 (0.012)**	0.165 (0.015)**	0.213 (0.020)**
Db223/229 (0.490) /227 (0.275)	0.064 (0.018)*	-0.039 (0.016)	0.073 (0.021)*	0.090 (0.028)*
Db411/244 (0.648) /242 (0.133)	-0.101 (0.022)**	0.081 (0.019)*	-0.127 (0.024)**	-0.160 (0.033)**
Db493/165 (0.351) /171 (0.223)	0.048 (0.009)**	-0.056 (0.008)**	0.051 (0.011)**	0.062 (0.014)*
Db681/255 (0.446) /275 (0.290)	-0.074 (0.017)*	0.034 (0.015)	-0.087 (0.019)**	-0.116 (0.025)**
	0.106 (0.026)*	-0.083 (0.023)*	0.124 (0.029)*	0.157 (0.040)*

\* $P < 0.05$ , \*\* $P < 0.01$ , \*\*\* $P < 0.001$ .<sup>a</sup>The two alleles at highest frequency over all localities.<sup>b</sup>Frequencies of these two alleles.

With fewer than 20 loci, the Wilcoxon test is the most powerful (Piry *et al.*, 1999), indicating significant heterozygosity deficiencies for MUB, GRD and TAM. In contrast to the BOTTLENECK results, the M analyses (Table 5) indicate significant bottlenecks for all of the Australian populations except ISG and GRD, with strongest evidence for the three southern populations (GER, MAL and BUL). Both BOTTLENECK and M methods assume no selection, but it is not known how the apparent selection at some loci (see below) would affect these results.

#### Geographical variation in genetic diversity and allele frequency

Expected heterozygosity ( $H_e$ ) decreased with increasing latitude ( $P < 0.01$ ). In the multiple regression analyses, all

four variables were significant for expected heterozygosity at the Db090, Db142 and Db681 loci (Table 6). The residuals from the models at these three loci showed no significant departure from normality. Allele frequencies at Db090, Db142, Db223, Db411, Db493 and Db681 showed significant effects generally for all four variables (Table 6). Residuals from the models for Db090, Db142 and Db411 did not depart from normality, but those at the other three loci did show significant departures. The significance levels for these three loci may be unreliable and should be interpreted cautiously. In all cases, the effect of longitude was of opposite sign to the effects of latitude and the two principal components.

#### Multilocus simulation tests for selection

To apply the ln RH test, we averaged the heterozygosities ( $H'$ ) for the Australian and Argentinian populations for

**Table 7** Results of the ln RH tests for selection

Locus	H <sup>a</sup>		ln RH <sup>b</sup>	ln RH <sup>c</sup>
	Australia	Argentina		
Db003	0.609	0.787	-1.202	-1.309
Db013	0.505	0.915	-3.698	-3.872
Db034	0.620	0.484	0.444	0.768
Db052	0.547	0.684	-0.791	-0.875
Db087	0.873	0.935	-1.091	-1.246
Db090	0.811	0.834	-0.142	-0.220
Db109	0.682	0.888	-1.793	-2.070
Db122	0.390	0.567	-0.808	-0.959
Db142	0.607	0.881	-2.320	-2.528
Db223	0.644	0.617	0.114	0.156
Db225	0.589	0.860	-2.258	-2.312
Db290	0.574	0.922	-3.648	-3.650
Db411	0.548	0.848	-2.207	-2.360
Db493	0.717	0.429	1.864*	1.766*
Db681	0.685	0.869	-1.727	-1.813
Mean			-1.284	-1.368
S.d.			1.498	1.555
Upper limit			1.652	1.679
Lower limit			-4.220	-4.416

\* $P < 0.05$ .

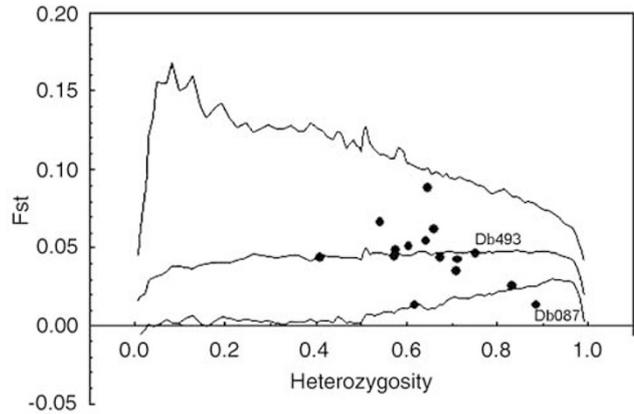
<sup>a</sup>Expected heterozygosity corrected for small sample size, and averaged over populations.

<sup>b</sup>ln (mean Australian  $H'$ /mean Argentine  $H'$ ), where  $H' = (1/(1-H))^2 - 1$ .

<sup>c</sup>Average (ln (pair-wise Australian  $H'$ /pair-wise Argentine  $H'$ )).

each locus, and used these averages to compute ln RH for each locus. The distribution of ln RH values under the assumption of neutrality is expected to be Gaussian (Kauer *et al.*, 2003). Thus, to determine significance of the ln RH values, the mean and standard deviation of the empirical ln RH distribution were used, with loci significant if their ln RH value falls outside the 95% limits ( $\pm 1.96$  s.d.). Of the 15 loci, 1 (Db493) was significant (ln RH = 1.864,  $P = 0.018$ ; Table 7). With only 15 loci, outliers may occur by chance, but accepting that the other 14 loci are neutral, and testing ln RH for Db493 against the normal distribution defined by these 14, the probability for Db493 is 0.004. Further, ln RH values and their distributions were computed for each locus for all 18 pair-wise comparisons of Australian and Argentinian populations. Although not an independent test, Db493 again was significant (ln RH = 1.766,  $P = 0.022$ ; Table 7). Further, 7 of the 18 comparisons for Db493 had ln RH values outside the 95% limits. The probability of observing seven significant tests for a given locus, calculated from the binomial distribution, is  $P = 1.5E-5$ . We note that Db493 showed more variation in Australian than in Argentinian populations, that is, not a selective sweep, but increased differentiation of the Australian populations.

The Beaumont test was run separately for the nine Australian populations, and for these plus the two Argentinian populations. In both cases, Db087 showed significantly less variation than expected, falling below the 0.025 quantile limit (Figure 5 for the nine Australian populations). For the two sets of populations, the test statistics and probability values were (-3.432, 0.0004) and (-6.711, 0.0000). With only 15 points, one might be expected to be outside the 95% limits. However, for both



**Figure 5**  $F_{ST}$  values estimated for 15 microsatellite loci in nine Australian populations plotted against heterozygosity, with 0.975, 0.50 and 0.025 quantiles of the conditional distribution, estimated from an island model with expected  $F_{ST} = 0.047$ .

sets, Db087 is well below the 0.005 quantile. Thus, there is strong evidence that Db087 is affected by forces that are different from those affecting the remaining loci, namely balancing selection that keeps alleles at similar frequencies in each of the populations.

## Discussion

Our primary aims were to assess the apparent bottleneck at colonization of Australia and to determine if any of the microsatellite loci marked genome regions that were either subject to selection or affected by selective sweeps. The studied Australian populations, which span 14.5° of latitude and range from the coast to 345 km inland, cover much of the range of *D. buzzatii* in Australia.

### Demographic history of *D. buzzatii* in Australia

Previous comparisons of Australian populations with original ones in Argentina support a bottleneck at the time of introduction, with the former showing reduced variability for allozyme (Barker *et al.*, 1985), mitochondrial DNA (Halliburton and Barker, 1993), microsatellite (Frydenberg *et al.*, 2002; Table 1) and other molecular (Picinali *et al.*, 2007) polymorphisms. Distributions of the overall Australian and Argentinian allele frequencies at 15 microsatellite loci (Figure 4) show not only the reduced variability of the former, but also dramatic shifts at many of the loci—clear signals of a founding bottleneck. Using the method of Ramirez *et al.* (2006), which makes no explicit assumption other than that sample sizes are adequate, we estimate that this was a moderate bottleneck of about 30–40 founders on colonization of Australia.

The BOTTLENECK and M ratio analyses gave different signals regarding past bottlenecks. Bottleneck failed to detect any evidence of a past bottleneck in any Australian population. However, this test is effective only for identifying populations that have recently experienced a severe reduction in population size, that is,  $< 4 N_e$  generations ago, where  $N_e$  is the bottleneck effective size (Piry *et al.*, 1999). As the colonization bottleneck in Australia occurred some 600–700 generations ago (assuming about 10 generations per year since

introduction in the period 1931–1936), and the bottleneck effective size was  $\leq 30$ –40 estimated number of founders, the failure to detect a past bottleneck is not unexpected. In fact, this test showed all populations to have a heterozygosity deficiency (significant for three of the six northern populations) suggesting that they are not at mutation–drift equilibrium, but that there has been a recent expansion in population size or a recent influx of rare alleles from genetically distinct immigrants. Given the relatively low (albeit significant) genetic differentiation among these populations, the latter possibility is most unlikely. However, the former is very compatible with the prediction (Barker, 1982; Sokal *et al.*, 1987) of a population explosion of *D. buzzatii* immediately following its introduction in the 1930s.

Garza and Williamson (2001) argued that their M ratio should retain information about past demographic history for longer than methods such as BOTTLENECK, and our results appear to validate this. In addition, Garza and Williamson (2001) show that the M ratio test is suited to the demographic scenario expected for *D. buzzatii*, that is, a severe, single-generation (or just a small number of generations) reduction in population size, followed by a rapid recovery. The M ratio test gave clear evidence of a past bottleneck for all except two of the Australian populations, with strongest evidence for the three southern populations.

Population differentiation was highly significant for each of the 15 loci and overall (Table 4). Pair-wise estimates show that this differentiation is much higher for the three southern populations as compared with the other six, and high among these three. The three southern localities are outside the predicted core distribution range for *D. buzzatii*, as determined by climatic suitability (Barker *et al.*, 2005), and the pair-wise  $F_{ST}$  estimates, the estimated number of founders and the reduced mean number of alleles per locus (Table 1) all suggest a secondary bottleneck for the southern populations.

The demographic history of *D. buzzatii* in Australia can be summarized as a bottleneck at colonization by founders from Argentina, rapid spread and expansion to an enormous population size over the area of the *Opuntia* infestation in Queensland and New South Wales, and then contraction to spatially isolated populations. Further colonizations within Australia with moderate secondary bottlenecks gave rise to the southern populations.

#### Tests for selection

Given the evidence for a bottleneck at colonization, we need to ensure that any evidence suggesting selection is not simply a function of the bottleneck. That is, can we distinguish effects of the founding bottleneck from later local adaptation to different environments after colonization? Clearly, the founding bottleneck changed allele frequencies in Australia as compared with the source population of Argentina (Figure 4). The populations studied here all derive from one original very large population, and the significant genetic differentiation among them may have been due to genetic drift. The significant isolation by distance and latitudinal clinal patterns might be consistent with drift, but the significant associations for some loci with climatic variables

after correcting for geographical location suggest selective differentiation, that is, local adaptation after colonization. On the other hand, as latitude and longitude were significant after controlling for climate effects (PC1 and PC2), geographical variation in factors other than climate also apparently is affecting allele frequencies. As the climate variables, summarized as PC1 and PC2, refer to long-term averages, other factors may include extremes that occur sporadically, for example, stressful conditions, such as drought or heat waves (episodic selection, Gillespie, 1991), and climatic effects summarized in other principal components.

Although the frequency of the standard second chromosome arrangement decreases with increasing latitude in Australia (Knibb *et al.*, 1987; Knibb and Barker, 1988) and in Argentina (Hasson *et al.*, 1995), indicating selective differentiation of populations for inversion frequencies, only Db034 and Db052 are included within an inversion (Figure 2), and these loci showed no significant effects in any selection analysis. Although Db142 is near the proximal breakpoint of inversion 4s, this inversion has not been found in Australian populations (Knibb *et al.*, 1987).

Although we had expected that some LD generated by the bottleneck at colonization may have been maintained, particularly for chromosome 2 loci within or near the inversion breakpoints, the finding of extensive LD between loci on different chromosomes was not predicted. Such LD is expected to be halved in each generation, and eliminated in about six generations. Thus, it could not have been generated by the small founding population, and subsequently maintained for 600–700 generations, unless due to epistatic selection. But if so, more consistency over populations in the expression of LD would be expected. Alternatively, the observed LD between loci on different chromosomes may be a transitory function of drift due to small  $N_e$  in the few generations before sampling. Again, however, more consistency would be expected, in this case over locus pairs within populations, with populations differing in the extent of LD. The BAR population, with 15.24% of significant pair-wise comparisons, may fit this explanation, as rots were very sparse at the time of collection, and much more collection effort was necessary at this site. Further, sampling of related individuals can lead to LD, a possibility for *D. buzzatii* as individual rots are founded by about 10 flies or less (Prout and Barker, 1989). Although a weak test, comparisons of observed and expected heterozygosities do not support this explanation.

Given this, we must question whether the apparent selection detected for seven loci also could be due to drift. For the loci where apparent selection was detected in the regression analyses, drift cannot be excluded, and further evidence will be necessary to demonstrate a causal relationship with any climatic variables. However, two of these loci (Db411 and Db493) are on chromosome 2. Of the 23 significant pair-wise LD among chromosome 2 loci, 14 were between one of the two loci within the inversion (Db034 and Db052) and one of the other five loci, suggesting selection on the inversions (standard vs  $2j$ ;  $2jz^3$  is rare in Australian populations; Knibb *et al.*, 1987), in addition to effects of drift. This inversion selection might then account for the apparent selection found for Db411 and Db493, and the significant

In RH test for Db493. However, the similar frequencies of the four common alleles of Db493 (Figure 4), but with each allele at highest frequency in a different population (Supplementary Table S2), are unlikely to have been generated by selection on the inversions or by drift. Db493 apparently is exhibiting local adaptation to the novel environments in Australia—such as, different species of the *Opuntia* host plants as compared with those in Argentina, and a broad range of climatic environments. Certainly, drift does not account for all genetic differentiation among these populations, as local adaptation has been shown for traits relevant for thermal adaptation (Sarup *et al.*, 2006).

The Beaumont test identified another locus (Db087), which showed a different signature of selection: balancing selection keeping alleles at similar frequencies in different populations, including those in Argentina. This test assumes that the loci are unlinked or loosely linked. Although Db087 showed significant LD with other loci on chromosome 4—with Db142 in one population and with Db681 in two populations—these loci are far distant from Db087 (Figure 2), and the significant LD seem unlikely to bias the test. As with the LD between loci on different chromosomes, these LD on chromosome 4 may result from drift. But again, more consistency might be expected over locus pairs within populations. Apparently Db087 is marking selection over a substantial proportion of chromosome 4.

Our finding that different approaches identify different loci is not surprising. Balancing selection at Db087 could not be detected by the other two approaches. Loci showing geographically varying selection (in the multiple regression analyses) will not be detected by the Schlötterer In RH test, unless like Db493, they also show more variation in the derived (Australian) than in the ancestral populations, and may well not be detected by the Beaumont test (showing higher than expected  $F_{ST}$  for a given heterozygosity, but not significant).

Linkage disequilibrium, expressed as the percentage of  $D'$  values that were significant, is higher in the Australian populations than in the Argentine CAT population, which is likely representative of the source population(s) of the Australian founders. Thus, there is good evidence for LD generation in Australia by the bottleneck at colonization and/or by subsequent drift or selection. The one problematical case relates to Db223 and Db225, which are contiguous in chromosomal localization, but show significant LD in only one population (BAR), and Db223 allele frequencies show significant location–climate associations whereas Db225 does not. However, of the loci that show significant associations with location–climate effects, those for Db223 are weakest, and regressions including PC1 or PC2 with latitude and longitude showed significant effects of longitude for both Db223 and Db225. Selective effects at these juxtaposed loci must be considered tentative.

In addition to drift effects, directional selection in the Australian populations would further promote LD (Maynard Smith and Haigh, 1974; Thomson, 1977; Kaplan *et al.*, 1989). Significant linkage disequilibria between loci that are not closely linked have been found in other species that have been subject to bottlenecks

and/or strong selection (Ledig *et al.*, 1999; Yan *et al.*, 1999; Kohn *et al.*, 2000; Sharbel *et al.*, 2000; McRae *et al.*, 2002; Sinervo and Clobert, 2003; Hansson *et al.*, 2004). Although LD is expected to decay at a rate dependent on the recombination fraction, and to be maintained only for very closely linked loci, these results show that linkage disequilibria may be far more common in natural populations than is generally assumed, and the loci apparently affected by selection may well be marking selection in large genome regions including many loci that are not necessarily closely linked.

A more complete understanding of the effects observed here will be possible only when individuals sampled are assayed both for inversion type and the microsatellite genotypes, so that LD between inversions and microsatellite loci can be estimated, and when both colonized populations and the endemic populations of Argentina are studied.

Hadrill *et al.* (2005) compared noncoding DNA polymorphism at 10 X-linked loci in three African and two non-African populations of *D. melanogaster*. Although several features of the data rejected the neutral model, it was concluded that simple bottleneck models were sufficient to account for most, if not all, polymorphism features. However, the demographic history of *D. melanogaster* is not well known, and this conclusion depends on assumptions made regarding the times of colonization from Africa. *D. buzzatii* would provide a useful complementary model to *D. melanogaster* for studies of demography and selection, given the known and much shorter time since colonization, and well-defined demographic history.

## Acknowledgements

We thank Cino Pertoldi for assistance in running the M program, to Malcolm Schug and two anonymous reviewers for helpful comments on an earlier version, to the Danish Natural Sciences Research Council for financial support through a center grant and a frame grant to VL, and to the Ministerio de Educación y Ciencia (Spain), for partial funding—grant BFU2005-02237 awarded to AR.

## References

- Archie JW (1985). Statistical analysis of heterozygosity data: independent sample comparisons. *Evolution* **39**: 623–637.
- Barker JSF (1982). Population genetics of *Opuntia* breeding *Drosophila* in Australia. In: Barker JSF and Starmer WT (eds). *Ecological Genetics and Evolution: The Cactus–Yeast–Drosophila Model System*. Academic Press: Sydney, Australia. pp 209–224.
- Barker JSF, Krebs RA, Davies HI (2005). Geographical distributions, relative abundance and coexistence of *Drosophila aldrichi* and *Drosophila buzzatii* in Australia. *Austral Ecol* **30**: 546–557.
- Barker JSF, Mulley JC (1976). Isozyme variation in natural populations of *Drosophila buzzatii*. *Evolution* **30**: 213–233.
- Barker JSF, Sene Fd, East PD, Pereira MAQR (1985). Allozyme and chromosomal polymorphism of *Drosophila buzzatii* in Brazil and Argentina. *Genetica* **67**: 161–170.
- Beaumont MA, Nichols RA (1996). Evaluating loci for use in the genetic analysis of population structure. *Proc R Soc Lond B Biol Sci* **263**: 1619–1626.

- Cavalli-Sforza LL (1966). Population structure and human evolution. *Proc R Soc Lond B Biol Sci* **164**: 362–379.
- Cornuet JM, Luikart G (1996). Description and power analysis of two tests for detecting recent population bottlenecks from allele frequency data. *Genetics* **144**: 2001–2014.
- Dodd AP (1940). *The Biological Campaign against Prickly Pear*. AH Tucker, Government Printer: Brisbane, Australia.
- Ender JA (1986). *Natural Selection in the Wild*. Princeton University Press: Princeton, NJ, USA.
- Frydenberg J, Pertoldi C, Dahlgaard J, Loeschcke V (2002). Genetic variation in original and colonizing *Drosophila buzzatii* populations analysed by microsatellite loci isolated with a new PCR screening method. *Mol Ecol* **11**: 181–190.
- Garza JC, Williamson EG (2001). Detection of reduction in population size using data from microsatellite loci. *Mol Ecol* **10**: 305–318.
- Gillespie JH (1991). *The Causes of Molecular Evolution*. Oxford University Press: New York.
- González J, Betran E, Ashburner M, Ruiz A (2000). Molecular organization of the *Drosophila melanogaster* Adh chromosomal region in *D. repleta* and *D. buzzatii*, two distantly related species of the *Drosophila* subgenus. *Chromosome Res* **8**: 375–385.
- González J, Nefedov M, Bosdet I, Casals F, Calvete O, Delprat A et al. (2005). A BAC-based physical map of the *Drosophila buzzatii* genome. *Genome Res* **15**: 885–892.
- Goudet J (2001) FSTAT, a program to estimate and test gene diversities and fixation indices, <http://www.unil.ch/izea/software/fstat.html>.
- Haddrill PR, Thornton KR, Charlesworth B, Andolfatto P (2005). Multilocus patterns of nucleotide variability and the demographic and selection history of *Drosophila melanogaster* populations. *Genome Res* **15**: 790–799.
- Halliburton R, Barker JSF (1993). Lack of mitochondrial DNA variation in Australian *Drosophila buzzatii*. *Mol Biol Evol* **10**: 484–487.
- Hansson B, Westerdahl H, Hasselquist D, Åkesson M, Bensch S (2004). Does linkage disequilibrium generate heterozygosity–fitness correlations in great reed warblers? *Evolution* **58**: 870–879.
- Hasson E, Rodriguez C, Fanara JJ, Naveira H, Reig OA, Fontdevila A (1995). The evolutionary history of *Drosophila buzzatii* XXVI. Macrogeographic patterns in the inversion polymorphism in New World populations. *J Evol Biol* **8**: 369–384.
- Hochberg Y (1988). A sharper Bonferroni procedure for multiple tests of significance. *Biometrika* **75**: 800–802.
- Houlder DJ, Hutchinson MF, Nix HA, McMahon JP (2000) ANUCLIM User Guide, Version 5.1. Canberra, Centre for Resource and Environmental Studies, Australian National University.
- Kaplan NL, Hudson RR, Langley CH (1989). The ‘hitchhiking effect’ revisited. *Genetics* **123**: 887–899.
- Kauer MO, Dieringer D, Schlötterer C (2003). A microsatellite variability screen for positive selection associated with the ‘Out of Africa’ habitat expansion of *Drosophila melanogaster*. *Genetics* **165**: 1137–1148.
- Kimmel M, Chakraborty R, King JP, Bamshad M, Watkins WS, Jorde LB (1998). Signatures of population expansion in microsatellite repeat data. *Genetics* **148**: 1921–1930.
- Knibb WR, Barker JSF (1988). Polymorphic inversion and esterase loci complex on chromosome 2 of *Drosophila buzzatii*. 2. Spatial variation. *Aust J Biol Sci* **41**: 239–246.
- Knibb WR, East PD, Barker JSF (1987). Polymorphic inversion and esterase loci complex on chromosome 2 of *Drosophila buzzatii*. 1. Linkage disequilibria. *Aust J Biol Sci* **40**: 257–269.
- Kohn MH, Pelz HJ, Wayne RK (2000). Natural selection mapping of the warfarin-resistance gene. *Proc Natl Acad Sci USA* **97**: 7911–7915.
- Ledig FT, Conkle MT, Bermejo-Velazquez B, Eguiluz-Piedra T, Hodgskiss PD, Johnson DR et al. (1999). Evidence for an extreme bottleneck in a rare Mexican pinyon: genetic diversity, disequilibrium, and the mating system in *Pinus maximartinezii*. *Evolution* **53**: 91–99.
- Lessios HA (1992). Testing electrophoretic data for agreement with Hardy–Weinberg expectations. *Mar Biol* **112**: 517–523.
- Liu K, Muse SV (2005). PowerMarker: an integrated analysis environment for genetic marker analysis. *Bioinformatics* **21**: 2128–2129.
- Mann J (1970). *Cacti Naturalized in Australia and their Control*. SG Reid, Government Printer: Brisbane.
- Mantel N (1967). Detection of disease clustering and a generalized regression approach. *Cancer Res* **27**: 209–220.
- Maynard Smith J, Haigh J (1974). Hitch-hiking effect of a favorable gene. *Genet Res* **23**: 23–35.
- McRae AF, McEwan JC, Dodds KG, Wilson T, Crawford AM, Slate J (2002). Linkage disequilibrium in domestic sheep. *Genetics* **160**: 1113–1122.
- Piccinali RV, Mascord LJ, Barker JSF, Oakeshott JG, Hasson E (2007). Molecular population genetics of the alpha-esterase5 gene locus in original and colonized populations of *Drosophila buzzatii* and its sibling *Drosophila koepferae*. *J Mol Evol* **64**: 158–170.
- Piry S, Alapetite A, Cornuet JM, Paetkau D, Baudouin L, Estoup A (2004). GENECLASS2: a software for genetic assignment and first-generation migrant detection. *J Hered* **95**: 536–539.
- Piry S, Luikart G, Cornuet JM (1999). BOTTLENECK: a computer program for detecting recent reductions in the effective population size using allele frequency data. *J Hered* **90**: 502–503.
- Prout T, Barker JSF (1989). Ecological aspects of the heritability of body size in *Drosophila buzzatii*. *Genetics* **123**: 803–813.
- R Development Core Team (2003) R Foundation for Statistical Computing: Vienna, Austria <http://www.R-project.org>, ISBN 3-900051-00-3.
- Ramirez O, Altet L, Ensenat C, Vila C, Sanchez A, Ruiz A (2006). Genetic assessment of the Iberian wolf *Canis lupus signatus* captive breeding program. *Conserv Genet* **7**: 861–878.
- Ranz JM, González J, Casals F, Ruiz A (2003). Low occurrence of gene transposition events during the evolution of the genus *Drosophila*. *Evolution* **57**: 1325–1335.
- Raymond M, Rousset F (1995). GENEPOP (version-1.2): population genetics software for exact tests and ecumenicism. *J Hered* **86**: 248–249.
- Rousset F (1997). Genetic differentiation and estimation of gene flow from F-statistics under isolation by distance. *Genetics* **145**: 1219–1228.
- Sarup P, Sørensen JG, Dimitrov K, Barker JSF, Loeschcke V (2006). Climatic adaptation of *Drosophila buzzatii* populations in southeast Australia. *Heredity* **96**: 479–486.
- SAS Institute (1985). *SAS User’s Guide: Statistics*. SAS Institute Inc.: Cary, North Carolina.
- Schlötterer C (2002). A microsatellite-based multilocus screen for the identification of local selective sweeps. *Genetics* **160**: 753–763.
- Schug MD, Smith SG, Tozier-Pearce A, McEvey SF (2007). The genetic structure of *Drosophila ananassae* populations from Asia, Australia and Samoa. *Genetics* **175**: 1429–1440.
- Sharbel TF, Haubold B, Mitchell-Olds T (2000). Genetic isolation by distance in *Arabidopsis thaliana*: biogeography and postglacial colonization of Europe. *Mol Ecol* **9**: 2109–2118.
- Sinervo B, Clobert J (2003). Morphs, dispersal behavior, genetic similarity, and the evolution of cooperation. *Science* **300**: 1949–1951.
- Sokal RR, Oden NL, Barker JSF (1987). Spatial structure in *Drosophila buzzatii* populations: simple and directional spatial autocorrelation. *Am Nat* **129**: 122–142.

- Sokal RR, Rohlf FJ (1981). *Biometry*. WH Freeman: San Francisco.
- Storz JF (2005). Using genome scans of DNA polymorphism to infer adaptive population divergence. *Mol Ecol* **14**: 671–688.
- Thomson G (1977). Effect of a selected locus on linked neutral loci. *Genetics* **85**: 753–788.
- Vilela CR, Sene Fd, Pereira MAQR (1980). On the *Drosophila* fauna of the Chaco and east slopes of the Andes in Argentina. *Rev Bras Biol* **40**: 837–841.
- Weir BS, Cockerham CC (1984). Estimating F-statistics for the analysis of population structure. *Evolution* **38**: 1358–1370.
- Wright S (1951). The genetical structure of populations. *Ann Eugen (London)* **15**: 323–354.
- Yan G, Romero-Severson J, Walton M, Chadee DD, Severson DW (1999). Population genetics of the yellow fever mosquito in Trinidad: comparisons of amplified fragment length polymorphism (AFLP) and restriction fragment length polymorphism (RFLP) markers. *Mol Ecol* **8**: 951–963.

Supplementary Information accompanies the paper on Heredity website (<http://www.nature.com/hdy>)