RESEARCH ARTICLE Open Acc

Morphological differentiation despite gene flow in an endangered grasshopper

Eddy J Dowle[1,2*], Mary Morgan-Richards[1] and Steven A Trewick[1]

Abstract

Background:

between populations are gradually accumulating, aided by increasingly sophisticated genetic tools [24-30].

Historically, one of the most informative animal groups in this field of study have been Orthoptera and in particular grasshoppers [14,31-35]. Here we report on flightless New Zealand short-horned grasshoppers (Orthoptera:

correspond to current taxonomic groups or morphological types. MtDNA sequences from

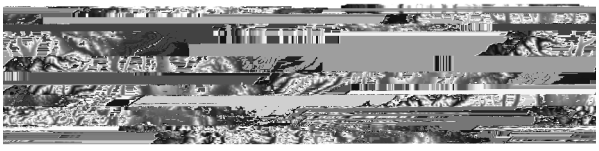
australis from either the central group $F_{CT} = 0.09175$ ($P = 0.09677$), or from the area of sympatry $F_{CT} = 0$ ($P = 1$) (Figure 1).

Microsatellites

The three microsatellite loci surveyed each had between 16 and 18 alleles. No evidence of linkage disequilibrium was detected and Hardy-Weinberg expectations were met in the majority of population samples. A positive relation-

i

ii



Nuclear sequencing

We amplified and sequenced the ITS region (706 bp including 5.8S, ITS1 and 2) from 40 grasshoppers. Some of the grasshoppers had unambiguous single ITS sequence but many had more than one ITS sequence, consistent with these grasshoppers being heterozygotes of mixed ancestry. Of twenty-five grasshoppers (15 *S. childi* and 10 *S. australis*) collected near the township of Alexandra, 16 (11 *S. childi*, 5 *S. australis*) had more than one sequence which differed by the presence of an INDEL approximately 100 bp from the ITS1 forward primer. Sequences of ITS2 from these individuals were unambiguous except at single nucleotide polymorphic sites (SNPs), confirming that these grasshoppers carried more than one ITS sequence per genome. There were 16 SNPs in the set of unambiguous sequences. However the presence of an INDEL near the start of ITS1 meant grasshoppers with more than one sequence had only 13 observable SNPs. Only one of the

Rad-Seq SNPs

The Illumina sequencing provided 9,789,323 forward reads of 100 bp for the 30 grasshoppers, of which 8,934,377 were retained after quality checks in process_radtags.pl (part of the Stacks package). These comprised between 3,743 and 978,246 reads per individual with a total of 30,439 loci. Three individuals (2 *S. australis* and 1 *S. childi*) were removed due to low coverage. From these data sets we identified 8,958 loci that occurred in >2 individuals and these were subjected to further selection based upon coverage per population. The Lindis population was removed from subsequent analysis as several individuals failed to produce data of sufficient quality (likely due to poor DNA quality) and most of the putative-loci resolved were not represented in the other samples. Of 8,958 loci, 74 were retained as they occurred in 50% of the two populations (*Sigaus childi* and *Sigaus australis* in sympatry). The relatively low proportion of loci that were represented across both population samples was due to insufficient representation of their very large genome despite our protocol involving quanitification and compensation for this. Nevertheless we obtained ample data for our purpose.

The distribution of pairwise (*S. childi*/*S. australis*) F_{ST} values for each of the 74 putative-loci revealed the high frequency of low scores expected in the absence of significant structure (Figure 6a). A test for deviations from expected frequencies of neutral loci in BayeScan indicated that one marker may have been subject to diversifying selection, log(PO) >0 alpha 0.878 (Figure 6b). A BLAST search of the sequence containing this SNP did not result in any matches to known sequences on Genbank. Mean population pairwise F_{ST} was low (0.025), with a confidence interval that effectively included zero (CI 2.5% 0.001, CI 97.5% 0.053), providing little evidence that these samples represent more than one population, with random mating. Population differentiation estimated with STRUCTURE suggested extensive sharing of genetic material among populations, with no species structure detected (Figure 4ci). This was confirmed by analysis in MIGRATE-N, which indicated extensive gene flow between the species in the zone of sympatry (Table 2).

Discussion

We found morphological support for two distinct entities consistent with their existing taxonomic treatment as species *Sigaus australis* and *Sigaus childi*. *Sigaus australis* has a comparatively wide geographic range that can be subdivided into a number of phylogeographically distinct mtDNA haplogroups. *Sigaus childi* is nested within *S. australis* in terms of habitat, geographic range and genetic diversity; in stark contrast to its clear morphological distinction. Despite occuring in sympatry, no phenotypic intermediates were detected. We found no evidence of genetic partitioning in putatively-neutral mtDNA sequence, microsatellite, ITS sequence loci or SNP data. None of the mtDNA diversity detected within *Sigaus australis* is concordant with current taxonomic subdivision. Microsatellite allele frequencies within *S. childi* are indistinguishable from those within sympatric *S. australis* suggesting recent (and on-going) gene flow. The SNP data show no population structure and extensive gene flow between the two stisti915.1()]T3S241i17.785Ts8147 0e0

[5,23]. In the Sigaus

using traditional species characteristics as the sole morphological traits analysed. Much of the taxonomy in the Sigausgenus relies on the pronotum shape, but descriptions are often vague, based on discrete states and inferred from few individuals making species identification difficult [36,65,66]. Using two digital images of the pronotum of each of 147 individuals (113 *S. australis*, 34 *S. childi*) that were obtained with the aid of a dissecting microscope we tested whether shape variation could be detected from metric data. Using IMAGEJ [68], 14 landmarks were identified around the perimeter of the dorsal surface of the pronotum on each image of each grasshopper and measured. The landmarks were selected to maximise variation among individuals. These measurements were analysed using MORPHOJ [69]. A procrustes fit aligned by principal axes was performed to eliminate size differences before a Procrustes ANOVA was used to examine the error of image capture. This analysis revealed that the error arising from image capture variation was biologically irrelevant: mean squares for image capture was 32 times smaller than the variation found between individual grasshoppers. Juveniles, adults and both sexes were included in the analyses and tested to confirm they did not

australis samples only, secondly among all samples from all areas sampled for the complex, and thirdly among all samples collected in the area of sympatry (*S. childi* and *S. australis*) and central group (*S. australis*) (Figure 1). The analyses were run using an admixture model with correlated allele frequency, 100,000 generations of burn-in followed by 100,000 generations, and the number of groups (K) set from 1 to 20 (10 replicates each). The optimum value of K was found using the K method except for K = 1, which was determined by examination of the bar-plots and structure harvester [81]. Charts were averaged over the 10 replicates and re-drawn using CLUMPP and distruct [82,83]. We sought evidence of genetic differentiation concordant with morphology using the populations within the central group subset identified by STRUCTURE (Figure 1).

A standard AMOVA was used to test for significant genetic differences based on the estimate of genetic partitioning among groups (F_{CT}) using ARLEQUIN version 3.5.1.2 [78]. The first run tested *S. childi* against all populations in the central group and the second only those *S. childi* and *S. australis* individuals from the area of sympatry (Figure 1).

Nuclear sequence

Nuclear sequences representing the internal transcribed spacers (ITS1 and ITS2) of the rRNA cluster and the intervening rRNA 5.8S gene were obtained using the primers ITS4 and ITS5 [84]. PCR conditions and sequencing followed standard protocols as above. Sequences were aligned using GENEIOUS PRO version 5.3.4 [74] and checked by eye. Sequences were generated for all grasshoppers from the area of sympatry of *S. childi* and *S. australis* (the Alexandra region). Alignment and comparison of unambiguous with ambiguous sequences allowed us to identify the most likely combinations of sequences that gave the observed heterozygotes [30]. Where sequence variants differed by single nucleotide substitutions we could identify and resolve the polymorphism. Where sequence variation involved INDELs the resulting length polymorphism was evident by abrupt onset of sustained nucleotide ambiguity

putative-locus (read), and each putative-locus appeared in both S. australis and S. childi populations, and occured in 50% of the individuals.

Gene flow between the two populations in sympatry was estimated using MIGRATE-N version 3.5.1 [92,93], although algorithms that test for gene flow are often not ideal for situations where gene flow is very high, which

17. Dowling TE, Secor CL: The role of hybridization and introgression in the diversification of animals. Annu Rev Ecol Syst 1997, 28:593–619.
18. Servedio MR, Noor MAF: The role of reinforcement in speciation: theory and data. Annu Rev Ecol Evol Syst 2003, 34:339–364.
19. Barton N, Hewitt G: Hybrid Zones and Speciation. In Evolution and speciation: essays in honor of MJD White; 1981:109.
20. Nosil P, Schluter D: The genes underlying the process of speciation. Trends Ecol Evol 2011, 26

73. Rambaut A Se-Al: Sequence Alignment Editor. v2.0a1 Oxford, UK: University of Oxford; 2002. Available from http://evolve.200.ox.ac.uk/.
74. Geneious version v5.3.4 created by Biomatters Available from http://www.geneious.com.
75. Bandelt H-J, Forster P, Röhl A Median-joining networks for inferring intraspecific phylogenies Mol Biol Evol 1999,16:37–48.
76. Mantel N The detection of disease clustering and a generalized regression approach Cancer Res 1967,27:209–220.
77. Jensen JL, Bohonak AJ, Kelley ST Isolation by Distance, web Service. BMC Genetics 6: 13. v.3.16 http; 2005. http://ibdws.sdsu.edu.
78. Excoffier L, Laval G, Schneider S Arlequin ver. 3.0: an integrated software package for population genetics data analysis Evol Bioinformatics Online 2005,1:47–50.
79. Van Oosterhout C, Hutchinson WF, Wills DPM, Shipley P Micro-Checker: software for identifying and correcting genotyping errors in microsatellite data Mol Ecol Notes 2004,4(3):535–538.
80. Carlsson J Effects of microsatellite null alleles on assignment testing. J Hered 2008,99(6):616–623.
81. Evanno G, Regnaut S, Goudet J Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Mol Ecol 2005,14:2611–2620.
82. Rosenberg NA Distruct: a program for the graphical display of population structucro-c(Ros7656:)-308.99e656:2.7(6 Tf 1ol)]TJ /F3 1 Tf 6.0394 0 T22008,99(6):61613Tj /F10 1 218]TJ /F2 1 Tf 2480 7.4999 5Jakobs984.6(V0.)T 262