Invaders weather the weather: rapid adaptation to a novel environment occurs despite reduced genetic diversity

Daniel Selechnik [1,2*], Mark F. Richardson [3,4], Richard Shine [1], Jayna DeVore [1], Simon Ducatez [1], and Lee A. Rollins [2]

[1] School of Life and Environmental Sciences (SOLES), University of Sydney 2006
[2] Evolution & Ecology Research Centre, School of Biological, Earth and Environmental Sciences, University of New South Wales, Sydney, NSW 2052
[3] Deakin Genomics Centre, School of Life and Environmental Sciences, Deakin University, Locked Bag 20000, Geelong, VIC, Australia 3216
[4] Centre for Integrative Ecology, School of Life and Environmental Sciences, Deakin University, Locked Bag 20000, Geelong, VIC, Australia 3216
[*] Correspondence: Dan Selechnik, Phone: +61 452 304 787; E-mail: danselechnik@gmail.com

Abstract

Invasive species often exhibit rapid evolution in their introduced ranges despite the genetic bottlenecks that are thought to accompany the translocation of small numbers of founders; however, some invasions may not fit this "genetic paradox." The invasive cane toad (*Rhinella marina*) displays high phenotypic variation across its environmentally heterogeneous introduced Australian range. Here, we used three genome-wide datasets to characterize population structure and genetic diversity in invasive toads: RNA-Seq data generated from spleens sampled from the toads' native range in French Guiana, the introduced population in Hawai'i that was the source of Australian founders, and Australia; RNA-Seq data generated from brains sampled more extensively in Hawai'i and Australia; and previously published RADSeq data from transects across Australia. We found that toads form three genetic clusters: (1) native range toads, (2) toads from the source population in Hawai'i and long-established areas near introduction sites in Australia, and (3) toads from more recently established northern Australian sites. In addition to strong divergence between native and invasive populations, we find evidence for a reduction in genetic diversity after introduction. However, we do not see this reduction in loci putatively under selection, suggesting that genetic diversity may have been maintained at ecologically relevant traits, or that mutation rates were high enough to maintain adaptive potential. Nonetheless, cane toads encounter novel environmental challenges in Australia and appear to respond to selection across environmental breaks; the transition between genetic clusters occurs at a point along the invasion transect where temperature rises and rainfall decreases. We identify loci known to be involved in resistance to heat and dehydration that show evidence of selection in Australian toads. Despite well-known predictions regarding genetic drift and spatial sorting during invasion, this study highlights that natural selection occurs rapidly and plays a vital role in shaping the structure of invasive populations.

Author Summary

Despite longstanding evidence for the link between genetic diversity and population viability, the "genetic paradox" concept reflects the observation that invasive populations are successful in novel environments despite a putative reduction in genetic diversity. However, some recent studies have suggested that successful invasions may often occur due to an absence of obstacles such as genetic diversity loss or novel adaptive challenges. The recent emergence of genome-wide technologies provides us with the tools to study this question comprehensively by assessing both overall genetic diversity, and diversity of loci that underlie ecologically relevant traits. The invasive cane toad is a useful model because there is abundant phenotypic evidence of rapid adaptation during invasion. Our results suggest strong genetic divergence between native and invasive populations, and a reduction in overall genetic diversity; however, we do not see this reduction when solely assessing ecologically relevant loci. This could be for reasons that support or refute the genetic paradox. Further studies may provide perspectives from other systems, allowing us to explore how variables such as propagule size affect the fit of an invasion to the model of the paradox. Studying invasive species remains important due to their largely negative impacts on the environment and economy.

Introduction

The genetic paradox of invasion [1] describes a phenomenon that challenges widespread evidence of the relationship between genetic diversity and adaptive potential. High genetic diversity within a population is beneficial because it likely underlies phenotypic variation across individuals, allowing the population to respond to selection imposed by environmental change [2, 3]. Furthermore, a greater number of alleles confers an increased frequency of heterozygosity, which is often associated with population fitness [2]. Small or isolated populations with low genetic diversity have been shown to suffer declines due to inbreeding depression and the associated reduction of individual fitness [4-6]. Conservation efforts to salvage such populations by introducing conspecific individuals from allopatric populations (thereby introducing new alleles into the population; "genetic rescue") have been successful, suggesting that the maintenance of genetic diversity can be crucial for population viability [5, 6].

Despite the fact that invasive populations are thought to undergo genetic bottlenecks (due to the translocation of a small number of founders from their native range to an introduced range [1, 7]), invasive species are also characterized by their ability to establish and spread in their introduced ranges. Invasion success is commonly linked to rapid evolution, including adaptation to novel environmental conditions over short timescales [8-11]. Additionally, some invaders exhibit novel phenotypic traits that enhance invasive potential, such as increased growth and dispersal rates [11, 12]. There are many examples of evolutionary change during invasion without high levels of genetic diversity [13].

Although low genetic diversity may limit the ability of an invasive population to respond to natural selection, rapid evolution can also occur through non-adaptive processes: (1) genetic drift may occur on range edges, reducing genetic diversity across an introduced range [14]; (2) due to spatial sorting, the invasion front may be inhabited exclusively by the

individuals with the highest dispersal rates (even if they are not the fittest) because they have arrived first and can only breed with each other [15], resulting in a geographic separation of phenotypes [15, 16]; (3) admixture or hybridization may occur between individuals from different introductions or sources [17].

It has recently been suggested that the genetic paradox of invasion may be rare; some invasive populations do not suffer a reduction in genetic diversity during introduction, and others do not face novel adaptive challenges in their introduced ranges [18]. Furthermore, some invasive systems demonstrate a 'spurious' paradox due to inadequate estimation of genetic diversity (i.e. too few markers), or to maintenance of genetic diversity only at ecologically relevant traits, or to a reduction in genetic diversity resulting from natural selection rather than from genetic bottlenecks [18]. These ideas can be tested with genome-wide data [19, 20].

The Australian cane toad (*Rhinella marina*) is an invader that has exhibited rapid evolution despite apparently low genetic diversity [21]. Cane toads were serially translocated from South America to Martinique and Barbados, to Jamaica, to Puerto Rico, to Hawai'i before 101 founders were introduced from the Hawai'ian island of Oahu to Queensland (QLD), Australia in 1935 [22]. Toads have since spread westward through the Northern Territory (NT) into Western Australia (WA; Fig 1). The native range and source area are tropical environments with relatively high mean annual rainfall (3200 mm in French Guiana, 1200-2000 mm in Hawai'i) and high temperatures (26°C in French Guiana, 22-24°C in Hawaii) [23]. The Australian range is heterogeneous in several environmental factors, particularly aridity; on average, sites in QLD receive more annual rainfall than do those in the NT and WA (2000-3000 mm in QLD, 400-1000 mm in NT and WA; Fig S1A), and have lower annual mean temperatures (21-24°C in QLD, 24-27°C in NT and WA; Fig S1B) [24]. Heritable phenotypic differences in behavior [25], thermal performance [26], morphology

[27, 28], and immune function [29] have been documented in cane toads at different localities across the Australian range. Nonetheless, surveys have reported low levels of genetic diversity, based on microsatellites [30] and MHC [31], and only one mitochondrial haplotype in the *ND3* gene of 31 individuals sequenced from Hawai'i and Australia [32].

Here, we assessed genetic diversity and population structure using single nucleotide polymorphisms (SNPs) identified in RNA-Seq data from samples originating from French Guiana (the "native range"), Hawai'i (the "source"), QLD (the "range core"), NT ("intermediate" areas), and WA (the "invasion front") to test whether the paradox is supported in the cane toad invasion, and to investigate the evolutionary dynamics during introduction. We predicted high divergence between native and invasive populations due to a combination of genetic drift, selection, and spatial sorting, but little genetic differentiation within invasion phases due to a putative lack of standing genetic diversity. Analyses to identify loci with outlier F_{ST} values or associations with environmental variables allow us to test for selection; because of the increase in aridity at intermediate areas and the invasion front, we predicted that loci under selection would be in genes involved in thermal tolerance, which may underlie the success that toads have had dispersing through these areas.

RNA-Seq data provide genome-wide information, but are limited to transcribed regions of the genome [33]. RNA-Seq data may therefore contain variants that exhibit strong signals for both demography and selection. Conversely, reduced representation sequencing approaches such as restriction site-associated DNA sequencing (RADSeq) data provide genome-wide information on a subset of all coding and non-coding sequences (only SNPs near restriction enzyme cleavage sites are detected); because non-coding sequences make up most of the genome, SNPs detected using this technique are more likely to represent neutral variation [34]. To focus on how selection may shape genomic diversity in invasive toads, we compared the results from our RNA-Seq data to those from a publicly available RADSeq

dataset [35] that we reanalyzed (the purpose of that study was not to analyze population structure per se, so we modified the analysis for our questions). The RADSeq dataset was limited to samples from QLD (the "range core" towards "intermediate" areas) and the border of NT/WA ("intermediate" areas towards the "invasion front").

Materials and Methods

Sample collection, RNA extraction, and sequencing

In November of 2017, we collected samples from French Guiana (native range; Fig 1; Table S1). In June of 2015, we collected samples from the Hawai'ian Islands (source; Fig 1; Table S1). In April and May of 2013-2015, we collected samples along an invasion transect in Australia (range core, intermediate areas, and the invasion front; Fig 1; Table S1). Upon collection, we excised whole spleen tissue (sample sizes in Table S1A) and whole brain tissue (sample sizes in Table S1B) from female toads immediately after euthanasia. Each tissue sample was initially preserved in RNAlater buffer (QIAGEN, USA), initially kept at -20ºC during fieldwork (less than one month), and then drained and transferred to a -80°C freezer for long-term storage.

We carried out RNA extractions using the RNeasy Lipid Tissue Mini Kit (QIAGEN, USA) following the manufacturer's instructions, with an additional genomic DNA removal step using on-column RNase-free DNase treatment (QIAGEN, USA). We quantified the total RNA extracted using a Qubit RNA HS assay on a Qubit 3.0 fluorometer (Life Technologies, USA). Extracts were then stored at -80◦C until sequencing. Sequencing was conducted commercially at Macrogen (Macrogen Inc., ROK). mRNA libraries were constructed using the TruSeq mRNA v2 sample kit (Illumina Inc., USA), which included a 300bp selection step. We sequenced 46 spleen samples and 72 brain samples in multiple batches (not pooled) across 3 lanes of Illumina HiSeq 2500. Capture of mRNA was performed using the oligo dT

method, and size selection parameter choices were made according to the HiSeq2500 manufacturer's protocol. Each individual sample was ligated with a unique barcode. Overall, this generated 709 million paired-end 2 x 125-bp reads from spleen data and 1.74 billion paired-end 2 x 125-bp reads from brain data. Raw sequence reads are available as FASTQ files in the NCBI short read archive (SRA) under the BioProject Accession PRJNA510261 (spleen data from French Guiana and Hawai'i), PRJNA395127 (spleen data from Australia), and PRJNA479937 (all brain data).

Data pre-processing and alignment

First, we examined per base raw sequence read quality (Phred scores) and GC content, and checked for the presence of adapter sequences for each sample using FastQC v0.11.5 [36]. We then processed raw reads (FASTQ format) from each sample with Trimmomatic v0.35 [37], using the following parameters: ILLUMINACLIP:TruSeq3-PE.fa:2:30:10:4 SLIDINGWINDOW:5:30 AVGQUAL:30 MINLEN:36. This removed any adaptor sequences, trimmed any set of five contiguous bases with an average Phred score below 30, and removed any read with an average Phred score below 30 or sequence length below 36 bp.

As a reference, we used the annotated *R. marina* transcriptome [38], which was constructed from brain, spleen, muscle, liver, ovary, testes, and tadpole tissue. We conducted per sample alignments of reads (FASTQ files) to the reference using STAR v2.5.0a [39] in basic two-pass mode with default parameters, a runRNGseed of 777, and specifying binary alignment map (BAM) alignment outputs. As STAR-generated BAM files lack read groups (identifiers for reads that specify the individual that they come from and the platform that was used to sequence them), we added them to our BAM files using the AddOrReplaceReadGroups tool in Picard Tools [40]. To avoid making incorrect variant calls, we removed duplicate reads (reads that map to the same position of the same gene in the

same individual) using the MarkDuplicates tool in Picard Tools. To split reads containing an N (region of the reference that is skipped in a read) into individual exon segments, we used the SplitNCigarReads tool in the Genome Analysis Toolkit (GATK) v3.8.0 [41].

Variant calling and filtering

To call SNPs and insertion-deletions (indels), we used the HaplotypeCaller tool in GATK [41] on our alignment (BAM) files. This tool works by identifying regions of the reference that show evidence of variation (called 'active regions'). Variants required a minimum phred-scaled confidence of 20 to be called (marked as passing the quality filters) and emitted (reported in the output), using the stand_call_conf and stand_emit_conf options.

To ensure that SNPs were called as accurately as possible, we performed this in 'GVCF' mode (which generates one intermediate genomic variant call format, or 'gVCF' file per individual sample). By using this mode, we avoided missing SNPs at loci that match the reference in some but not all individuals. We then used the GenotypeGVCFs tool to merge the gVCF files, re-calculate genotype likelihoods at each SNP locus across all individuals, and re-genotype and re-annotate all SNP loci. The results were written to one merged VCF file. Although we initially genotyped all spleens and brains together, we discovered during downstream analyses that there was an effect of tissue type on population assignment – even from the same individuals, and using only SNPs from transcripts expressed in both tissues, spleen and brain samples were assigned to separate populations). Thus, we genotyped spleens and brains separately, resulting in two merged VCF files, and subsequently kept them separate for all downstream analyses. We retained both datasets because each provides a unique benefit: the spleen dataset includes native range samples, but the brain dataset has more extensive sampling of the invasive populations.

Rather than following a random distribution, some SNPs are clustered. Clustered SNPs may not be independent, and are thus filtered. We used the VariantFiltration tool to identify and filter ‘clusters’ (sets of 3 SNPs that appear within a window of 35 bases) in each of our merged VCF files. We also used this tool to filter variants with QualByDepth (QD; variant confidence divided by the unfiltered depth of non-reference samples) less than 2.0, depth of coverage (DP) less than 20.0, and allele frequency less than 0.05. We then subset our VCF files to include only the variants that passed the filters we set in the VariantFiltration step using the SelectVariants tool. This resulted in 803,489 SNPs from spleen data and 818,536 SNPs from brain data. We used bcftools [42] to further filter the VCF files, only retaining biallelic SNPs. We examined the results of filtering for minimum minor allele frequency (min MAF) thresholds of 0.01 or 0.05, and several missing data tolerance (MDT) thresholds (the maximum percentage of individuals in the dataset in which a genotype for a locus can be absent without that locus being filtered out). Our population structure results were consistent across min MAF and MDT thresholds, so we ultimately chose to filter our data at min MAF = 0.05 and MDT = 0% (no missing data tolerated) because some downstream analyses cannot handle missing data. These filtering steps reduced the number of SNPs to 65,195 in spleen data and 35,842 in brain data (Table S1A-B).

Inference of population structure

We used PLINK [43] to convert our VCF files to the Browser Extensible Data (BED) format, which is readable to fastStructure. We then used fastStructure [44] to infer population structure using a variational Bayesian framework for calculating posterior distributions, and to identify the number of genetic clusters in our dataset (K) using heuristic scores [44]. We ran the structure.py ten times each (K= 1 to 10). We then took the resulting meanQ files from fastStructure and plotted them using the pophelper package [45] in R [46].

In addition to using fastStructure for population assignment, we also performed a Redundancy Analysis (RDA) using the vegan package [47] in R, which fits genetic and environmental data to a multivariate linear regression, and then performs a principal component analysis (PCA) on the fitted values. RDA visualizes both population structure and the effects that environmental variables may have in shaping it. To do this, we downloaded climatic data on French Guiana, Hawai'i, and Australia from the Bioclim database [48] using the raster package [23]. Because these areas vary in aridity, we downloaded data on rainfall and temperature; these data are averages of annual statistics over the period of 1970 to 2000. Specifically, we downloaded: annual mean temperature, maximum temperature of the warmest month, minimum temperature of the coldest month, annual precipitation, precipitation of the wettest quarter, and precipitation of the driest quarter. We then used the vcfR package [49] to convert our VCF files to the GENLIGHT format, which is readable to the vegan package.

Identification of candidate loci under selection

Loci that are under natural selection may have abnormally high F_{ST} values, causing them to be outliers among all other loci in the transcriptome. Bayescan v2.1 [50] detects loci with outlier F_{ST}s using the multinomial-Dirichlet model. We used PGDSpider [51] to convert our VCF files to the BAYESCAN format, and then ran Bayescan to detect loci with outlier pairwise F_{ST} values between our three genetic clusters in spleen data (native range toads versus source and core toads versus intermediate and frontal toads) and our two genetic clusters in brain data (source and core toads versus intermediate and frontal toads).

Loci under natural selection also may have allele frequencies that are associated with environmental variables. We used the climatic data that we downloaded from the Bioclim database [48] using the raster package [23] on rainfall during the driest quarter and maximum

temperature in the warmest month. We then used the lfmm v2.0 package [52] in R to perform a latent factor mixed model (LFMM) to test the association between the allele frequencies of every locus and these two environmental variables. We applied a Benjamini-Hochberg correction to all p-values from the LFMM.

Because these scans reveal thousands of loci putatively under selection, and because outlier tests have high false positive rates [53], we took a more conservative approach by cross-matching our list of F_{ST} outliers with each of our lists of environmentally-associated loci. For individual investigation, we only looked at loci with both an outlier F_{ST} value and an association with a putatively significant environmental variable.

Evaluation of genetic differentiation and diversity

To quantify levels of genetic differentiation and diversity, we computed basic statistics in the hierfstat [54] and diveRsity [55] packages in R [46]. We first used PGDSpider v2.1.1.3 [51] to convert our VCF files to the FSTAT and GENEPOP formats, which are readable to hierfstat and diveRsity, respectively. We then used these packages to calculate measures of genetic differentiation, including global F_{ST} and pairwise F_{ST} (by genetic cluster, invasion phase, and collection site), and genetic diversity, such as expected heterozygosity (He) and rarefied allelic richness (AR). We also computed Shannon's Information Index (SI) with the dartR package [56]. After calculating these measures of diversity across all loci (N=65,195 from spleen data, N=35,842 from brain data), we calculated the same measures in loci putatively under selection: those with outlier F_{ST} values (N=648 from spleen data, N=203 from brain data) and those associated with environmental variables (N=4179 from spleen data, N=530 from brain data). In spleen data, this allowed us to investigate the hypothesis that genetic diversity is maintained at ecologically relevant traits even if genome-wide diversity is lost. In brain data, this allowed us to examine the effect of natural selection on

genetic diversity within Hawai'i and Australia. We used Kruskal-Wallis tests to assess the significance of the differences in genetic diversity.

Annotation of SNPs

To visualize the types of genomic regions in which our SNPs lie, we used SnpEff [57]. SnpEff analyzes the information in a VCF file, annotates the SNPs, and estimates their effects. We calculated the relative proportions of each type of SNP for both tissue types using the full dataset of SNPs, and our two selection candidate datasets (containing F_{ST} outlier loci associated with an environmental variable). We then used the stats v3.5.0 R package to perform z-tests on the differences in proportions of SNP types between the full and selection candidate datasets. We expected differences in the proportions of certain types of SNPs, such as a lower proportion of synonymous variants in the selection candidates because synonymous mutations do not result in a codon change (and thus are unlikely to change protein function, i.e. phenotype).

Isolation by distance

To examine the effects of geographic distance on genetic distance across the Australian range, we performed a Mantel test using ade4 v1.7-5 [58]. Because we were focused on isolation by distance through range expansion after introduction (to Australia), samples from Hawai'i and French Guiana were excluded; and because we sampled brain tissue in a greater number of sites within Australia (as opposed to spleen tissue; Table S1A-B), we only performed the Mantel test on brain data. For the geographic distance matrix, we used the dist function in R [46] to calculate the Euclidean distances in geographic space between collection sites based on their coordinates. For the genetic distance matrix, we used the collection site-based pairwise F_{ST} values generated from hierfstat. We performed four Mantel

tests, using: (1) all loci, (2) loci with outlier F_{ST} values, (3) loci with outlier F_{ST} values and with an environmental association (temperature or rainfall), (4) loci with outlier F_{ST} values but without an environmental association. A significant result across all loci may suggest that the prominent driver of genetic structure is genetic drift. However, clinal variation in allele frequencies may also result from spatial sorting (particularly in loci underlying dispersal ability), or from selection associated with environmental clines; distinguishing these evolutionary processes is difficult. Thus, we performed the additional tests to attempt to separate the effects of genetic drift and spatial sorting. Loci with outlier F_{ST} values are candidates for natural selection and spatial sorting; thus, a significant result from only these loci would less likely be caused by genetic drift. Outlier loci can be further separated into those with an environmental association, which are likely candidates for natural selection driven by environmental variables, and those without an environmental association, which are possible candidates for spatial sorting. If spatial sorting is indeed influencing genetic differentiation, then we may expect to see different patterns between these two groups of loci (e.g. linearity only in the outlier loci without an environmental association).

Acquisition, processing, and analysis of RADSeq dataset

In addition to our own RNA-Seq dataset, we downloaded a publicly available RADSeq dataset [35] to evaluate population structure in Australian cane toads. Unlike RNA-Seq data, which consists of coding sequences only [34], RADSeq data include SNPs from across the entire genome, but only around sites that are cleaved by a selected set of restriction enzymes [33]. This method allows us to contrast the results of assessing loci potentially under ‘tighter’ selection (e.g., codon bias) with those primarily constituting neutral variation (RADSeq). Individuals in the RADSeq dataset were collected from QLD (range core towards intermediate areas, N=179) and the border of NT/WA (intermediate areas towards the

invasion front, N=441; Fig 1). However, samples were lacking from a large part of the range between those regions; thus, identifying geographic points of differentiation between genetic clusters may be unlikely using this dataset. Furthermore, although there were many collection sites within these regions, metadata with coordinates (or even location names) were unavailable. We downloaded raw 100-bp reads from NCBI short read archive (SRA) under the BioProject Accession PRJNA328156. We then converted SRA files to the FASTQ format using the fastq-dump tool in the SRA toolkit.

We used Stacks v2.0 [59] for all RADSeq data processing. First, we used the process_radtags program to remove low quality reads from the FASTQ files. Next, we used the denovo_map pipeline to perform a *de novo* assembly for each sample, align matching DNA regions across samples (called 'stacks'), and call SNPs using a maximum likelihood framework. We allowed a maximum of five base mismatches between stacks within an individual and three base mismatches between stacks between individuals. This resulted in 499,623 SNPs. We then filtered the results using the populations program, using one random SNP per locus and four combinations of parameter choices (min MAF=0.05, MDT = 0.5; min MAF = 0.05, MDT = 0.2; min MAF = 0.01, MDT = 0.5; min MAF = 0.01, MDT = 0.2). Structure results were consistent across min MAF values, but not MDT values: two distinct populations were only detected using the more stringent MDT choice. Because population structure in this dataset is relatively low, and because it is appropriate to use more stringent filtering choices when analyzing datasets with low population structure (Selechnik et al., unpublished; Chapter 2), we ultimately chose to filter this dataset with min MAF = 0.05 and MDT = 0.2. Furthermore, using these filtering choices produced results that were most consistent with those reported by Trumbo et al. (2016) and with those of our RNA-Seq dataset. Filtering reduced the number of SNPs to 8,296. The results were written to a VCF file and a STRUCTURE file (which could be read into fastStructure).

We ran fastStructure using the same methodology as we did for our RNA-Seq dataset. In addition to the model-based fastStructure, we also performed a discriminant analysis of principal components (DAPC) on this dataset using adegenet [60]. DAPC is a multivariate approach that identifies the number of genetic clusters using K-means of principal components and a Bayesian framework. We also converted the VCF to the FSTAT and GENEPOP formats and computed the same basic statistics for the RADSeq dataset as we did for our RNA-Seq datasets using hierfstat and diveRsity. Because collection site coordinate metadata were unavailable for the RADSeq dataset, we did not perform a Mantel test using these data.

Results

Population structure across the Hawai'ian-Australian invasion

We found three genetic clusters in our spleen data (Fig 2A): French Guiana (native range) formed its own genetic cluster, Hawai'i (source) clustered with QLD (core), and NT (intermediate) clustered with WA (invasion front). These patterns were mirrored by our brain data, which did not include native range individuals but had more extensive sampling in Hawai'i and Australia than did the spleen data. In our brain dataset, population differentiation seemed to align with environmental barriers (Fig 2B-C): Hawai'i (source) clustered with coastal QLD (core), whereas inland QLD/NT (intermediate) clustered with WA (invasion front). Inference of substructure within these two genetic clusters revealed that source and core toads may further differentiate into two separate groups (Fig 2D); however, intermediate and frontal toads remain one genetic cluster (Fig 2E).

RDA was mostly consistent with fastStructure for both spleen (Fig 3A) and brain (Fig 3B) data, except for the differentiation of source toads and core toads as two separate genetic clusters. It also suggested that: (1) Source toads may have diverged from toads of other

clusters due to selection from rainfall during the driest quarter. (2) Core toads may have diverged from toads of other clusters due to selection from rainfall during the wettest quarter, mean annual rainfall, and minimum temperature during the coolest month. (3) Intermediate and frontal toads may have diverged from toads of other clusters due to selection from maximum temperature during the hottest month and mean annual temperature.

Tests for selection

To identify loci putatively under selection, we searched for those with outlier F_{ST} values or associations with environmental variables. Out of 65,195 SNPs from spleen data (N=46 individuals across native and invasive populations), 648 were identified as F_{ST} outliers, with a mean global F_{ST} of 0.89. Mean pairwise F_{ST} values across outlier loci were 0.93 between the native and source/core populations, 0.97 between the native and intermediate/front populations, and 0.34 between the source/core and intermediate/front populations. From the same dataset (of 65,195 SNPs), 3088 SNPs were associated with maximum temperature during the warmest month and 1813 were associated with rainfall during the driest quarter. Of these loci, 772 were associated with both environmental variables. We focused on SNPs that were detected both as F_{ST} outliers and environmental correlates: there were 64 outlier SNPs associated with maximum temperature and 47 outlier SNPs associated with rainfall in the driest quarter (including 18 outliers associated with both environmental variables).

Most of the 64 outlier SNPs associated with maximum temperature during the hottest month (Table S2A) were in transcripts with functions such as metabolism, transcription regulation, and immune function. There were also two SNPs in *HSP4*, a gene encoding heat shock protein (HSP) 4 (70 kDa), which is involved in thermal tolerance (on exposure to heat stress) and protein folding and unfolding [61]. Most of the 47 outlier SNPs associated with

rainfall during the driest quarter (Table S3A) were in transcripts with functions such as cell signaling and immune function.

Out of 35,842 SNPs from brain data (N=72 individuals across invasive populations only), 203 were identified as F_{ST} outliers, with a mean global F_{ST} of 0.34. Mean pairwise F_{ST} was 0.50 between the source/core and intermediate/front populations. From the same dataset (of 35,842 SNPs), 345 SNPs were associated with maximum temperature during the warmest month and 194 were associated with rainfall during the driest quarter. Nine of these loci were associated with both environmental variables. Again, we focused on SNPs that were detected both as F_{ST} outliers and environmental correlates: there were 38 outlier SNPs associated with maximum temperature and 33 outlier SNPs associated with rainfall in the driest quarter; no outliers were associated with both environmental variables.

Most of the 38 outlier SNPs associated with maximum temperature during the hottest month (Table S2B) were in genes involved in cell signaling, mitochondrial processes (metabolism), or gene expression [61]. Similarly to the spleen data, there was one SNP from brain data in *HSP4*. Another SNP was in *ARNT2*, a gene encoding aryl hydrocarbon receptor nuclear translocator 2, a transcription factor involved in the hypothalamic pituitary adrenal (HPA) axis and visual and renal function [61]. Of the 33 outlier SNPs from brain data correlated with rainfall during the driest quarter (Table S3B), 19 were in *MAGED2*, a gene encoding melanoma-associated antigen D2, which is involved in renal sodium ion absorption [61]. Four were in *STXBP1*, a gene encoding syntaxin-binding protein 1, which is involved in platelet aggregation and vesicle fusion [61]. The rest were in genes generally involved in gene expression and cell signaling [61].

Evaluation of genetic differentiation, diversity, and isolation by distance

Our spleen data suggest that divergence from the native population is strong in the source/core population (pairwise F_{ST} = 0.29), and even stronger in the intermediate/frontal population (pairwise F_{ST} = 0.32). Differentiation between the source/core and intermediate/frontal populations is much lower (pairwise F_{ST} = 0.08). Overall divergence among natives and invaders is moderately high (global F_{ST} = 0.18)

These trends are consistent with our brain data, in which genetic differentiation between the source/core and intermediate/frontal populations is also low (pairwise F_{ST} = 0.07). In terms of substructure within the two invasive groups, differentiation between source and core toads was even lower (pairwise F_{ST} = 0.04 in brain data). Generally, pairwise F_{ST} by invasion phase (Table S4A) and collection site (Table S4B) from brain data revealed that toads from areas that are more closely linked by invasion history are less strongly differentiated. Because the brain data did not include native range samples, overall divergence was estimated to be much lower than it was in the spleen data (global F_{ST} = 0.04).

Diversity statistics from spleen data suggest that the native range population is only slightly more genetically diverse than the source/core population (native: He = 0.27, AR = 1.74, SI = 0.40; source/core: He = 0.26, AR = 1.74, SI = 0.40; p = 5E-4 for He). There is a greater drop in diversity from the native range population to the intermediate and frontal population (intermediate/front: He = 0.23, AR = 1.68, SI = 0.36; p < 2E-16 for all measures). However, we also estimated these measures by collection site (rather than by genetic cluster) to detect changes in genetic diversity at a finer scale; these calculations reveal a more obvious reduction in genetic diversity from the native population to all invasive collection sites (Table 1).

Table 1. Expected heterozygosity (He), allelic richness (AR), and Shannon's Information Index (SI) estimated using SNPs from spleen RNA-Seq data on native and invasive populations of cane toads (*Rhinella marina*) at collection sites across French Guiana, Hawai'i, and Australia.

Collection site	All loci (He, AR, SI)	Outlier F_{ST} loci (He, AR, SI)	Environmentally associated loci (He, AR, SI)
French Guiana (native)	0.27, 1.66, 0.40	0.03, 1.25, 0.06	0.17, 1.53, 0.25
Kapolei Park (source)	0.23, 1.60, 0.34	0.07, 1.20, 0.10	0.28, 1.77, 0.41
Haiku Gardens (source)	0.23, 1.60, 0.34	0.07, 1.22, 0.11	0.28, 1.75, 0.41
Gordonvale (core)	0.24, 1.63, 0.36	0.07, 1.21, 0.10	0.30, 1.84, 0.45
Daintree (core)	0.24, 1.62, 0.35	0.08, 1.24, 0.12	0.29, 1.79, 0.44
Cape Crawford (intermediate)	0.22, 1.59, 0.32	0.02, 1.06, 0.03	0.22, 1.59, 0.32
Timber Creek (intermediate)	0.21, 1.58, 0.32	0.02, 1.05, 0.03	0.22, 1.60, 0.32
Halls Creek (front)	0.21, 1.57, 0.32	0.02, 1.05, 0.03	0.19, 1.58, 0.30
Durack River (front)	0.22, 1.58, 0.32	0.03, 1.11, 0.05	0.20, 1.60, 0.31

Despite a slight reduction in genetic diversity from the native to invasive populations genome-wide, our spleen data suggest that diversity was not lost during introduction in subsets of loci putatively under selection. Among loci with outlier F_{ST} values, there is no reduction in genetic diversity from the native population to the source/core population; rather, it is higher in the source/core population (native: He = 0.03, AR = 1.25, SI = 0.06; source/core: He = 0.09, AR = 1.37, SI = 0.14; p < 2E-16 for all measures). However, diversity in these loci is reduced in the intermediate/front population (intermediate/front: He = 0.02, AR = 1.14, SI = 0.04). These trends are also shown in environmentally-associated loci (native: He = 0.17, AR = 1.53, SI = 0.25; source/core: He = 0.37, AR = 1.99, SI = 0.55; intermediate/front: He = 0.23, AR = 1.81, SI = 0.35). Calculations of these measures by collection site further support these findings (Table 1).

Similarly to the spleen data, genome-wide diversity statistics from brain data suggest that the source/core population is more diverse than the intermediate/frontal population (source/core: He = 0.34, AR = 2.00, SI = 0.51; intermediate/front: He = 0.31, AR = 1.97, SI = 0.47; p < 2E-16 for comparisons of all measures). Calculations by invasion phase (Table S5) suggest a reduction in diversity from core to invasion front. This reduction is greater in loci

putatively under selection: in those with outlier F_{ST} values, the source/core population is much more diverse than the intermediate/front population (source/core: He = 0.38, AR = 2.00, SI = 0.55; intermediate/front: He = 0.15, AR = 1.70, SI = 0.24; $p < 2E\text{-}16$ for comparisons of all measures). These trends are also shown in environmentally-associated loci (source/core: He = 0.35, AR = 2.00, SI = 0.53; intermediate/front: He = 0.08, AR = 1.57, SI = 0.15; $p < 2E\text{-}16$ for comparisons of all measures). Calculations of these measures by collection site further support these findings (Table 2).

Table 2. Expected heterozygosity (He), allelic richness (AR), and Shannon's Information Index (SI) estimated using SNPs from brain RNA-Seq data on native and invasive populations of cane toads (*Rhinella marina*) at collection sites across Hawai'i and Australia.

Collection site	All loci (He, AR, SI)	Outlier F_{ST} loci (He, AR, SI)	Environmentally associated loci (He, AR, SI)
Paradise Park (source)	0.28, 1.57, 0.41	0.31, 1.76, 0.45	0.30, 1.84, 0.45
Mauna Lani (source)	0.27, 1.56, 0.40	0.24, 1.61, 0.35	0.29, 1.73, 0.42
Kapolei Park (source)	0.29, 1.57, 0.43	0.29, 1.83, 0.43	0.25, 1.73, 0.38
Haiku Gardens (source)	0.30, 1.59, 0.44	0.31, 1.75, 0.45	0.31, 1.83, 0.46
Gordonvale (core)	0.30, 1.61, 0.44	0.36, 1.82, 0.51	0.22, 1.62, 0.33
Cairns (core)	0.32, 1.60, 0.48	0.38, 1.95, 0.55	0.24, 1.84, 0.38
Daintree (core)	0.29, 1.60, 0.43	0.32, 1.90, 0.48	0.18, 1.57, 0.28
Croydon (intermediate)	0.26, 1.60, 0.37	0.18, 1.45, 0.26	0.10, 1.24, 0.14
Burketown (intermediate)	0.28, 1.57, 0.41	0.18, 1.44, 0.26	0.09, 1.31, 0.14
Cape Crawford (intermediate)	0.28, 1.58, 0.42	0.15, 1.42, 0.23	0.08, 1.32, 0.13
Mataranka (intermediate)	0.28, 1.57, 0.41	0.16, 1.46, 0.24	0.07, 1.25, 0.11
Timber Creek (intermediate)	0.27, 1.55, 0.40	0.05, 1.18, 0.09	0.06, 1.22, 0.10
Purnululu (front)	0.29, 1.55, 0.44	0.20, 1.48, 0.12	0.07, 1.37, 0.11
Halls Creek (front)	0.27, 1.56, 0.40	0.09, 1.25, 0.14	0.06, 1.21, 0.10
Durack River (front)	0.27, 1.56, 0.41	0.13, 1.37, 0.19	0.06, 1.24, 0.10

Geographic distance in collection sites and pairwise F_{ST} values by collection site (as estimated with brain data) were significantly associated across all four of our Mantel tests using: (1) all loci ($p = 1E\text{-}3$, $r = 0.76$; Fig 4A); (2) all loci with outlier F_{ST} values ($p = 1E\text{-}3$, $r = 0.75$; Fig 4B); (3) only loci with outlier F_{ST} values and an environmental association ($p =$

1E-3, r = 0.62; Fig 4C); (4) only loci with outlier F_{ST} values but without an environmental association (p = 1E-3, r = 0.77; Fig 4D).

SNP Annotations

Annotation of SNPs revealed that most of our loci were either missense, synonymous, or found in the 3' untranslated region (UTR). A small percentage were found in the 5' UTR, and the remaining loci involved the loss or gain of stop or start codons (Figs S2-S3). In the spleen data, the temperature-associated outlier F_{ST} subset had a significantly higher proportion of synonymous variants (p = 8E-3) than the full set of SNPs. The outlier loci within *HSP4* were both synonymous variants. However, there were no significant differences in proportions between the rainfall-associated outlier F_{ST} subset and the full set of SNPs.

In the brain data, there were no significant differences in proportions of SNP variant types between the temperature-associated outlier F_{ST} subset and the full set. The locus within *HSP4* was a 3' UTR variant, and the locus within *ARNT2* was synonymous. The rainfall-associated outlier F_{ST} subset had a significantly higher proportion of 3' UTR variants (p = 5E-3) and lower proportion of missense variants (p = 0.04) than did the full set. Eighteen of the nineteen loci within *MAGED2* were 3' UTR variants (the last was missense), and two of the four loci within *STXBP1* were missense (the other two were synonymous).

Population structure in Australian toads from the RADSeq experiment

Analysis of the RADSeq dataset through fastStructure and DAPC also showed two genetic clusters within Australia (Fig 5A-B): QLD (core towards intermediate areas) in the first group, and the border of NT/WA (intermediate areas towards the invasion front) in the second group. There was little differentiation between these two groups (pairwise F_{ST} = 0.09) or overall (global F_{ST} = 0.04). We found no evidence of significant substructure within either

of the two genetic clusters. The two groups were equal in allelic richness (AR = 2.00 for both groups), but expected heterozygosity was higher in core toads than in intermediate/frontal toads (core: He = 0.50; intermediate/front: He = 0.33), which was consistent with our RNA-Seq results.

Discussion

We predicted that genetic diversity would be significantly lower in invasive populations than in native populations due to genetic bottlenecks taking place during the serial introductions of cane toads across Hawai'i and Australia. We also predicted strong divergence between native range and invasive populations due to a combination of genetic drift, selection and spatial sorting. In terms of population structure, we predicted low genetic structure within invasive populations in Hawai'i and Australia resulting from putatively low levels of standing genetic variation. Our analyses show evidence of a reduction in genetic diversity from native to invasive populations. We found that the toads form three genetic clusters: (1) native range toads form their own cluster, (2) Hawai'ian source toads cluster with eastern Australian range core toads, and (3) all toads from further west along the recent invasion transect cluster together (intermediate through to the invasion front). Differentiation is much higher between the native population and the invasive populations than between the invasive populations. The loci with the highest differentiation between native and invasive populations (outlier pairwise F_{ST} values > 0.9 between native and invasive) have relatively low diversity across all populations; this may be because these loci are homozygous in most individuals with no shared alleles between populations (fixed). Associations between our genetic structure patterns and allele frequencies with environmental variables suggest that natural selection is occurring in the western part of the Australian range. However, the significant result of our Mantel test across all loci also reveals gradual genetic drift across invasion phases; and the

significant result and high correlation coefficient of our Mantel test of F_{ST} outliers without environmental associations may be due to spatial sorting. It seems likely that all these forces are responsible for the observed population structure within invasive populations.

Our data suggest that cane toads may fit the model of the genetic paradox; we found evidence of strong divergence between native and invasive cane toad populations, and of a reduction in overall genetic diversity. Previous studies on mtDNA [32] and microsatellite [62] data have also suggested losses in genetic diversity from the native range to Hawai'i and Australia. Despite the reduction in genetic diversity from native to invasive populations across the full set of loci in the spleen data, we did not find this reduction in loci putatively under selection (loci with outlier F_{ST} values or associations with environmental variables) in the source/core population. This may provide support for the prediction that some invasions do not represent a genetic paradox despite overall loss of genetic diversity and the presence of novel adaptive challenges because genetic diversity at ecologically relevant traits is maintained by balancing selection [18]. However, diversity of this subset of loci is not just maintained in the source/core population; it is higher than in the native population. This may reflect admixture between toads from multiple introductions from different areas of the native range to Puerto Rico, where Hawai'ian toads were sourced from; the original South American sources for the Puerto Rican introductions included toads from both French Guiana and Guyana [22]. Alternatively, this may represent a genuine paradox; *de novo* mutation rates may have been high enough to restore or even enhance adaptive potential in the source/core population [18]; although these toads have been in Hawai'i and Australia for less than 100 years, they first diverged from native South American toads in the early 1800s when toads were introduced to the islands that would eventually (directly or indirectly) provide the source populations for the Puerto Rican introduction (i.e. Martinique and Barbados). From the source/core population to the intermediate frontal population, however, diversity of loci

putatively under selection is highly reduced; we believe this is because of directional selection, as outlined below.

Our two invasive genetic clusters (source/core and intermediate/front) diverge at the transition from coastal to inland areas within Australia, in an area where temperatures become higher and rainfall becomes scarcer, particularly during the dry season [24]. This correlation suggests that these environmental variables may drive population structure and that toads may be adapting to local conditions. Furthermore, RDA implicates selection from temperature and rainfall in driving divergence between invasive populations. Heterogeneous climates across introduced or expanding ranges have previously been linked to genetic differentiation, possibly due to climatically-imposed selection [10, 63].

Diversity estimates from our brain data suggest a reduction in diversity from the source/core population to the intermediate/frontal population in the full set of loci, but a much greater reduction in diversity in the subset of loci putatively under selection. The source (Hawai'i) and core (QLD) areas represent earlier phases of the invasion and experience similar environmental conditions to those of the native range, whereas intermediate (NT) and invasion front (WA) areas are more arid. Thus, the greater depletion in diversity at loci putatively under selection across the Australian invasion may be due to the effects of directional selection from the harsher climate at more recently colonized areas of the Australian range [18, 64].

Among our loci under selection associated with temperature are a few within a gene encoding protein HSP 4 (70 kDa). HSPs protect protein folding during increased temperatures and provide cells with enhanced thermal tolerance [65], and expression of HSP genes has been shown to underlie adaptive responses to environmental stress [66]. Changes in HSP levels in response to thermal environment manipulation vary between native and invasive cane toads, as well as between populations of toads within Australia [67]. In our

study, the SNPs within *HSP4* included two synonymous variants (spleen data) and a 3' UTR variant (brain data); 3' UTR variants are known to alter levels of protein functioning through differential expression of genes [68]. In the case of heat shock proteins, adaptation may not necessarily mean a change in function, but rather a change in the expression levels of proteins involved in the function, which could be caused by a 3' UTR variant. Another notable protein was aryl hydrocarbon receptor nuclear translocator 2 (from the *ARNT2* gene), a transcription factor that specifically acts on genes involved in the HPA axis and visual and renal function. The HPA axis is stimulated by high temperatures [69], and the secretion of corticosterone can be modulated adaptively in response to thermal change [70]. Elevation of corticosterone secretion in toads increases the rate of desiccation, thus making it maladaptive in arid environments [71]. Corticosterone levels in Brazilian and Australian toads in response to manipulation of thermal environment display similar patterns to those of HSPs [67]. Selection on loci potentially affecting corticosterone secretion by affecting the HPA is thus in line with these results. However, the SNP in *ARNT2* is synonymous; thus, this SNP might not change protein function.

More than half (19 of 33) of the loci under selection associated with rainfall (in brain data) were found in *MAGED2* (encodes melanoma-associated antigen D2). This protein regulates the expression and localization of two co-transporters that facilitate renal sodium ion absorption, which prevents excess loss of water and solutes via reabsorption through the kidneys (to be returned to the extracellular fluid and circulatory system) [72]. Similarly to the SNP within *HSP4*, 18 of 19 SNPs within *MAGED2* were 3' UTR variants, which can affect gene expression [68]. The high number of SNPs found in the 3' UTR of *MAGED2* is likely responsible for the difference in proportions of variant types between the full dataset and the rainfall-associated outlier dataset. Four more SNPs were found in *STXBP1* (encodes syntaxin-binding protein 1, which is involved in vesicle fusion and blood clotting). Dehydration is

known to increase blood clotting rate [73], and excessive blood clotting can be harmful to the heart, brain, and limbs [74]. Two of the four SNPs in *STXBP1* were missense variants, possibly producing change in protein function. Evolved changes in the rates of renal sodium absorption and blood clotting may allow toads from intermediate and invasion front areas to survive in drier environments. The remaining SNPs associated with rainfall or temperature were in genes generally involved in gene expression or signal transduction; these genes may facilitate expression of (or signaling to) the genes we have discussed, or may have independent functions in cellular maintenance.

Invasive species have been shown to rapidly adapt phenotypically to heterogeneous climatic conditions, allowing them to expand their ranges [75]. Some invaders tolerate higher temperatures [76] and water loss [77] better than do related indigenous taxa [78]. There is evidence of this in cane toads: under dehydrating conditions, wild-caught individuals from Australia exhibit better locomotor performance than do conspecifics from Hawai'i and the native range [79], and within Australia, toads from semi-arid areas (i.e. NT) exhibit better locomotor performance than do conspecifics from wetter areas (i.e. coastal QLD) [79, 80]. This trait is heritable; captive-bred toads with parental origins from hotter areas (northwestern Australia) outperform those with parental origins from cooler areas (northeastern Australia) at high (but not low) temperatures in a common-garden setting [26]. Coupled with the previously shown heritable phenotypic patterns, our results on genetic structure suggest that intermediate and frontal toads in Australia may be evolving enhanced thermal tolerance, thereby facilitating their continued westward range expansion [81]. However, experiments pairing genetic and phenotypic data from the same individuals are required to test this hypothesis.

Although our patterns of genetic structure (cluster separation along a distinct climatic gradient) and the identities of loci putatively under selection suggest that invasive toads are

responding to localized natural selection, our data also support the presence of genetic drift. The significant result of our Mantel test using all loci suggests that isolation by distance is occurring in the Australian toad population. Slight reductions in diversity (SI, AR, and He) along the invasion trajectory found here are expected and have been seen in other invasions [14]. However, the results of our additional Mantel tests suggest that another evolutionary force may also be at work: spatial sorting is the separation of individuals within a range-expanding population along phases of their expansion based on their dispersive capabilities [15], and may lead to clinal variation in dispersal-related allele frequencies along the range. Genetic differentiation in our loci with outlier F_{ST} values (which are candidates for natural selection and spatial sorting) is significantly associated with geographic distance. Furthermore, separating F_{ST} outlier loci with an environmental association (likely candidates for natural selection) from those without (possible candidates for spatial sorting) yields different patterns. Although the Mantel tests on both datasets are significant, the test on the environmentally associated outliers unsurprisingly shows two discrete groups of pairwise F_{ST} values – low pairwise F_{ST} values at low geographic distances and high pairwise F_{ST} values at high geographic distances (Fig 4C). Interestingly, the test on outliers without an environmental association shows a clinal increase in pairwise F_{ST} values with geographic distance (like those on the full dataset and the full outlier dataset), which may suggest that some of these loci are influenced by spatial sorting (Fig 4D).

Our results from the RADSeq dataset [35] were similar to those from our RNA-Seq dataset; both showed two genetic clusters within Australia. However, unlike the RNA-Seq results (which show a divide occurring in inland QLD, where temperature increases and rainfall decreases), the RADSeq results cluster all QLD toads in the first group, and NT/WA toads in the second group. In the RADSeq dataset, the spatial location of the transition between the two groups is difficult to pinpoint due to the absence of intermediate sampling

sites. The difference in group divisions between the two datasets may be because selection from environmental variables causes differentiation between toads from coastal and inland QLD (as seen in the RNA-Seq data), whereas demographic processes such as gene flow may lessen differentiation between them (as seen in the RADSeq data). Alternatively, toads from coastal and inland QLD may cluster together in the RADSeq data because the sampling sites in that study were south of those in our study, and the environmental conditions of coastal and inland sites are more similar in that area [48]. Estimates of diversity from the RADSeq dataset were slightly higher than those from the RNA-Seq dataset; this may be because neutral variation is expected to be higher than adaptive variation due to the effects of directional and purifying selection [64].

In conclusion, it appears that serial introductions have led to a reduction in genetic diversity of invasive cane toads. Nonetheless, cane toads do indeed face novel adaptive challenges in their introduced ranges, and are able to respond genetically to natural selection, particularly in the harsh environmental conditions of northwestern Australia. The ability of toads to adapt to these conditions may reflect the maintenance of diversity at ecologically relevant loci, or it may reflect sufficiently high mutation rates to bolster adaptive potential; thus, it remains unclear whether the toad invasion represents a true genetic paradox. Natural selection, genetic drift, and spatial sorting are the most plausible mechanisms by which genetic differentiation may occur, but require further investigation to tease apart. Determining the association between phenotypic traits and the occurrence of certain SNPs (such as the candidates that we have identified) at the individual level may be especially informative. Additionally, genetic changes may not be the only mechanisms by which adaptation occurs; studies on plasticity and epigenetic changes may also be useful for uncovering the mysteries of rapid evolution.

Acknowledgements

This work was supported by the Australian Research Council (FL120100074, DE150101393) and the Equity Trustees Charitable Foundation (Holsworth Wildlife Research Endowment). We thank Olivier Francois and Natalie Hofmeister for their input during discussions about our analyses. We thank Cam Hudson, Joachim Ehlenz, Serena Lam, and Chris Jolly for their assistance with sample collection. We thank Andrea West for her assistance during RNA extractions. All procedures involving live animals were approved by the University of Sydney Animal Care and Ethics Committee.

References

1. Allendorf FW. Introduction: Population Biology, Evolution, and Control of Invasive Species. Conservation Biology. 2003;17(1):24 - 30.
2. Reed DH, Frankham R. Correlation between Fitness and Genetic Diversity. Conservation Biology. 2003;17(1):230 - 7.
3. Frankham R. Genetics and extinction. Biological Conservation. 2005;126(2):131-40.
4. Blomqvist D, Pauliny A, Larsson M, Flodin LA. Trapped in the extinction vortex? Strong genetic effects in a declining vertebrate population. BMC Evolutionary Biology. 2010;10:33.
5. Westemeier RL, Brawn JD, Simpson SA, Esker TL, Jansen RW, Walk JW, et al. Tracking the long-term decline and recovery of an isolated population. Science. 1998;282:1695-8.
6. Madsen T, Shine R, Olsson M, Wittzell H. Restoration of an inbred adder population. Nature. 1999;402:34 - 5.
7. Barrett SCH, Kohn JR. Genetics and conservation of rare plants. New York: Oxford University Press; 1991.
8. Franks SJ, Munshi-South J. Go forth, evolve and prosper: the genetic basis of adaptive evolution in an invasive species. Molecular Ecology. 2014;23:2137-40.
9. Gao Y, Li S, Zhan A. Genome-wide single nucleotide polymorphisms (SNPs) for a model invasive ascidian Botryllus schlosseri. Genetica. 2018;146(2):227-34.
10. Leydet KP, Grupstra CGB, Coma R, Ribes M, Hellberg ME. Host-targeted RAD-Seq reveals genetic changes in the coral Oculina patagonica associated with range expansion along the Spanish Mediterranean coast. Mol Ecol. 2018;27(11):2529-43.
11. Whitney KD, Gabler CA. Rapid evolution in introduced species, 'invasive traits' and recipient communities: challenges for predicting invasive potential. Diversity and Distributions. 2008;14:569-80.
12. White TA, Perkins SE, Heckel G, Searle JB. Adaptive evolution during an ongoing range expansion: the invasive bank vole (Myodes glareolus) in Ireland. Mol Ecol. 2013;22(11):2971-85.
13. Rollins LA, Moles AT, Lam S. High genetic diversity is not essential for successful introduction. Ecology and Evolution. 2013;3:4501 - 17.

14. Rollins LA, Woolnough AP, Wilton AN, Sinclair R, Sherwin WB. Invasive species can't cover their tracks: using microsatellites to assist management of starling (Sturnus vulgaris) populations in Western Australia. Mol Ecol. 2009;18(8):1560-73.
15. Shine R, Brown GP, Phillips BL. An evolutionary process that assembles phenotypes through space rather than through time. Proceedings of the National Academy of Sciences of the United States of America. 2011;108:5708 - 11.
16. Hudson CM, Brown GP, Shine R. It is lonely at the front: contrasting evolutionary trajectories in male and female invaders. . Royal Society Open Science 2016;3:160687.
17. Mader G, Castro L, Bonatto SL, de Freitas L. Multiple introductions and gene flow in subtropical South American populations of the fireweed, Senecio madagascariensis(Asteraceae). Genet Mol Biol. 2016;39(1):135-44.
18. Estoup A, Ravigné V, Hufbauer R, Vitalis R, Gautier M, Facon B. Is There a Genetic Paradox of Biological Invasion? Annual Review of Ecology, Evolution, and Systematics. 2016;47:51-72.
19. Wellband KW, Pettitt-Wade H, Fisk AT, Heath DD. Standing genetic diversity and selection at functional gene loci are associated with differential invasion success in two non-native fish species. Mol Ecol. 2018;27(7):1572-85.
20. Willoughby JR, Harder AM, Tennessen JA, Scribner KT, Christie MR. Rapid genetic adaptation to a novel environment despite a genome-wide reduction in genetic diversity. Mol Ecol. 2018;27(20):4041-51.
21. Rollins LA, Richardson MF, Shine R. A genetic perspective on rapid evolution in cane toads (Rhinella marina). Mol Ecol. 2015;24(9):2264-76.
22. Turvey N. A Toad's Tale. Hot Topics from the Tropics. 2009;1:1 - 10.
23. Hijmans RJ. raster: Geographic data analysis and modeling. 2015.
24. Bureau of Meteorology AG. Climate Data Online: Commonwealth of Australia; 2018 [Available from: http://www.bom.gov.au/.
25. Gruber J, Brown GP, Whiting MJ, Shine R. Is the behavioural divergence between range-core and range-edge populations of cane toads (Rhinella marina) due to evolutionary change or developmental plasticity? Royal Society Open Science. 2017;4.
26. Kosmala GK, Brown GP, Christian KA, Hudson CM, Shine R. The thermal dependency of locomotor performance evolves rapidly within an invasive species Ecology and Evolution. 2018;8(9):4403-8.
27. Hudson CM, Brown GP, Shine R. It is lonely at the front: contrasting evolutionary trajectories in male and female invaders. R Soc Open Sci. 2016;3(12):160687.
28. Hudson CM, Brown GP, Stuart K, Shine R. Sexual and geographical divergence in head widths of invasive cane toads, Rhinella marina (Anura: Bufonidae), is driven by both rapid evolution and plasticity. Biological Journal of the Linnean Society. 2018;124:188-99.
29. Brown GP, Phillips BL, Dubey S, Shine R. Invader immunology: invasion history alters immune system function in cane toads (Rhinella marina) in tropical Australia. Ecol Lett. 2015c;18(1):57-65.
30. Leblois R, Rousset F, Tikel D, Moritz C, Estoup A. Absence of evidence for isolation by distance in an expanding cane toad (Bufo marinus) population: an individual-based analysis of microsatellite genotypes. Molecular Ecology. 2000;9:1905 - 9.
31. Lillie M, Shine R, Belov K. Characterisation of major histocompatibility complex class I in the Australian cane toad, Rhinella marina. PLoS One. 2014;9(8):e102824.
32. Slade RW, Moritz C. Phylogeography of Bufo marinus from its natural and introduced ranges. Proceedings of the Royal Society of London Series B: Biological Sciences. 1998;265:769.
33. Davey JW, Blaxter ML. RADSeq: next-generation population genetics. Briefings in Functional Genomics. 2011;9(5-6):416-23.

34. Wang Z, Gerstein M, Snyder M. RNA-Seq: a revolutionary tool for transcriptomics. Nature Reviews Genetics. 2009;10(57-63).
35. Trumbo DR, Epstein B, Hohenlohe PA, Alford RA, Schwarzkopf L, Storfer A. Mixed population genomics support for the central marginal hypothesis across the invasive range of the cane toad (Rhinella marina) in Australia. Molecular Ecology. 2016;25(17):4161–76
36. Andrews S. FastQC: a quality control tool for high throughput sequence data. 2010 [Available from: http://www.bioinformatics.babraham.ac.uk/projects/fastqc.
37. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 2014;30(15):2114-20.
38. Richardson MF, Sequeira F, Selechnik D, Carneiro M, Vallinoto M, Reid JG, et al. Improving amphibian genomic resources: a multi-tissue reference transcriptome of an iconic invader. GigaScience. 2018;7(1):1-7.
39. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: ultrafast universal RNA-seq aligner. Bioinformatics. 2013;29(1):15-21.
40. Institute B. Picard Tools. http://broadinstitute.github.io/picard. 2018.
41. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Research. 2010;20:1297-303.
42. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics. 2009;25(16):2078-9.
43. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, et al. PLINK: a toolset for whole-genome association and population-based linkage analysis. American Journal of Human Genetics. 2007;81.
44. Raj A, Stephens M, Pritchard JK. fastSTRUCTURE: Variational Inference of Population Structure in Large SNP Data Sets. Genetics. 2014;197(2):573 - 89.
45. Francis RM. pophelper: an R package and web app to analyse and visualize population structure Molecular Ecology Resources. 2016;17(1):27-32.
46. Team RC. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2016.
47. Oksanen J, Blanchet FG, Friendly M, Kindt R, Legendre P, McGlinn D, et al. vegan: Community Ecology Package. R package. 2018.
48. Hijmans RJ, Cameron SE, Parra JL, Jones PG, Jarvis A. Very high resolution interpolated climate surfaces for global land areas. International Journal of Climatology. 2005;25:1965–78.
49. Knaus BJ, Grünwald NJ. vcfr: a package to manipulate and visualize variant call format data in R. Molecular Ecology Resources. 2017;17(1):44-53.
50. Foll M, Gaggiotti O. A Genome-Scan Method to Identify Selected Loci Appropriate for Both Dominant and Codominant Markers: A Bayesian Perspective. Genetics. 2008;180:977-93.
51. Lischer HE, Excoffier L. PGDSpider: an automated data conversion tool for connecting population genetics and genomics programs. Bioinformatics. 2012;28(2):298-9.
52. Frichot E, Francois O. LEA: An R package for landscape and ecological association studies. Methods in Ecology and Evolution. 2015;6(8):925-9.
53. Whitlock MC, Lotterhos KE. Reliable Detection of Loci Responsible for Local Adaptation: Inference of a Null Model through Trimming the Distribution of FST. The American Naturalist. 2015;186(S1):24-36.
54. Goudet J. HIERFSTAT, a package for R to compute and test hierarchical F-statistics. Molecular Ecology Notes. 2005;5:184 - 6.

55. Keenan K, McGinnity P, Cross TF, Crozier WW, Prodohl P. diveRsity:AnR package for the estimation and exploration of popula tion genetics parameters and their associated errors. Methods in Ecology and Evolution. 2013;4:782-8.
56. Gruber B, Unmack PJ, Berry OF, Georges A. dartr: An r package to facilitate analysis of SNP data generated from reduced representation genome sequencing. Molecular Ecology Resources. 2018;18(3):691-9.
57. Cingolani P, Platts A, Wang LL, Coon M, Nguyen T, Wang L, et al. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff. Fly (Austin). 2012;6(2):80-92.
58. Thioulouse J, Dray S. Interactive Multivariate Data Analysis in R with the ade4 and ade4TkGUI Packages. Journal of Statistical Software. 2007;22(5):1 - 14.
59. Catchen J, Hohenlohe PA, Bassham S, Amores A, Cresko WA. Stacks: an analysis tool set for population genomics. Mol Ecol. 2013;22(11):3124-40.
60. Jombart T, Ahmed I. adegenet 1.3-1: new tools for the analysis of genome-wide SNP data. Bioinformatics. 2011;27(21):3070–1.
61. Consortium TU. UniProt: the universal protein knowledgebase. Nucleic Acids Res. 2017;45:D158-D69.
62. Estoup A, Wilson IJ, Sullivan C, Cornuet JM, Moritz C. Inferring Population History From Microsatellite and Enzyme Data in Serially Introduced Cane Toads, Bufo marinus. Genetics. 2001;159:1671-87.
63. Weber E, Schmid B. Latitudinal population differentiation in two species of Solidago (Asteraceae) introduced into Europe. American Journal of Botany. 1998;85(9):1110-21.
64. Andrews CA. Natural Selection, Genetic Drift, and Gene Flow Do Not Act in Isolation in Natural Populations. Nature Education Knowledge. 2010;3(10):5.
65. Kiang JG, Tsokos GC. Heat Shock Protein 70 kDa: Molecular Biology, Biochemistry, and Physiology. Pharmacology & Therapeutics. 1998;80(2):183 - 201.
66. Chen B, Feder ME, Kang L. Evolution of heat-shock protein expression underlying adaptive responses to environmental stress. Mol Ecol. 2018;27(15):3040-54.
67. Kosmala GK, Brown GP, Shine R. Kicking back Down Under: Invasive populations of cane toads (Rhinella marina) down-regulate their responses to thermal and hydric stress in a climatically harsh environment. Submitted. 2018.
68. Skeeles LE, Fleming JL, Mahler KL, Toland AE. The impact of 3'UTR variants on differential expression of candidate cancer susceptibility genes. PLOS ONE. 2013;8(3):e58609.
69. Malmkvist J, Damgaard BM, Pedersen LJ, Jorgensen E, Thodberg K, Chaloupkova H, et al. Effects of thermal environment on hypothalamic-pituitary-adrenal axis hormones, oxytocin, and behavioral activity in periparturient sows Journal of Animal Science. 2009;87(9):2796–805.
70. Telemeco RS, Addis EA. Temperature has species-specific effects on corticosterone in alligator lizards. Gen Comp Endocrinol. 2014;206:184-92.
71. Jessop TS, Letnic M, Webb JK, Dempster T. Adrenocortical stress responses influence an invasive vertebrate's fitness in an extreme environment. Proc Biol Sci. 2013;280(1768):20131444.
72. Greger R. Physiology of renal sodium transport. Am J Med Sci. 2000;319(1):51-62.
73. El-Sabban F, Fahim MA, Al Homsi MF, Singh S. Dehydration accelerates in vivo platelet aggregation in pial arterioles of lead-treated mice. J therm Biol. 1996;20(6):469-76.
74. PubmedHealth. Excessive Blood Clotting: NIH; 2014 [Available from: https://www.ncbi.nlm.nih.gov/pubmedhealth/PMH0062998/.
75. Colautti RI, Barrett SCH. Rapid Adaptation to Climate Facilitates Range Expansion of an Invasive Plant. Science. 2013;342(6156):364 - 6.

76. Braby CE, Somero GN. Following the heart: temperature and salinity effects on heart rate in native and invasive species of blue mussels (genus Mytilus). Journal of Experimental Biology. 2006;209:2554 - 66.
77. Godoy O, Lemos-Filho JP, Valladares F. Invasive species can handle higher leaf temperature under water stress than Mediterranean natives. Environmental and Experimental Botany. 2011;71(2):207 - 14.
78. Zerebecki RA, Sorte CJB. Temperature Tolerance and Stress Proteins as Mechanisms of Invasive Species Success. PLOS ONE. 2011;6(4):e14806.
79. Kosmala G, Christian K, Brown GP, Shine R. Locomotor performance of cane toads differs between native-range and invasive populations. The Royal Society Open Science. 2017;4:170517.
80. Tingley R, Greenlees MJ, Shine R. Hydric balance and locomotor performance of an anuran (Rhinella marina) invading the Australian arid zone. Oikos 2012;121:1959-65.
81. Szucs M, Vahsen ML, Melbourne BA, Hoover C, Weiss-Lehman C, Hufbauer RA. Rapid adaptive evolution in novel environments acts as an architect of population range expansion. Proc Natl Acad Sci U S A. 2017;114(51):13501-6.

Figures:

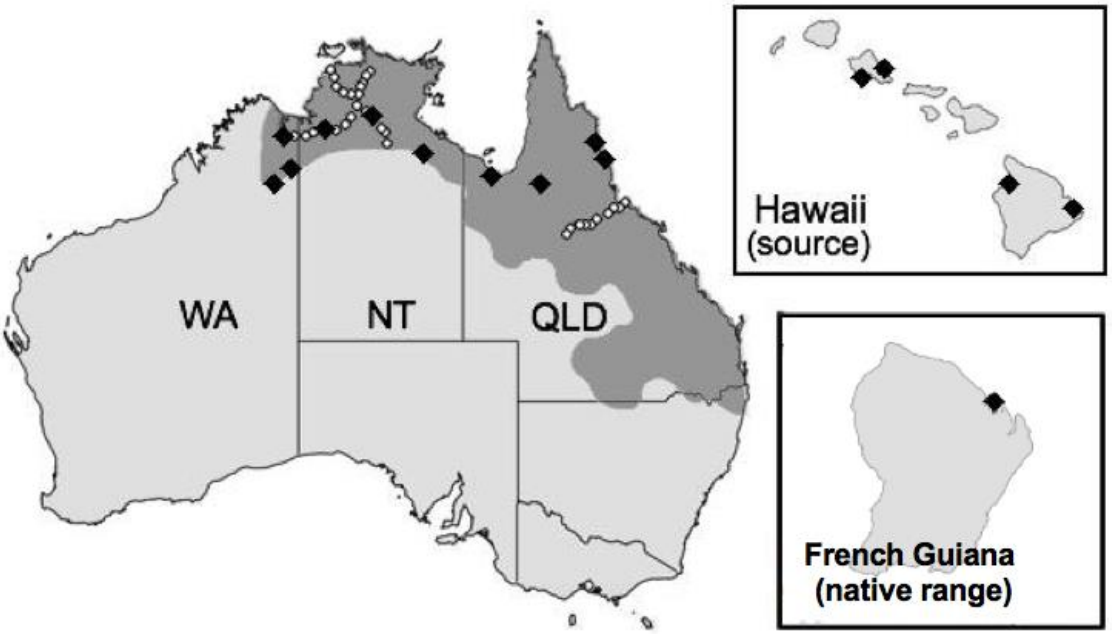

Fig 1. Map of two invaded territories (Australia and Hawai'i) and the native range (French Guiana) of the cane toad (*Rhinella marina*). The shaded region represents the toad's current Australian range, which is continuing to expand. Black diamonds in French Guiana, Hawai'i, and Australia indicate collection sites for our RNA-Seq experiment. White circles in Australia indicate collection sites for the RADSeq experiment from which we downloaded data.

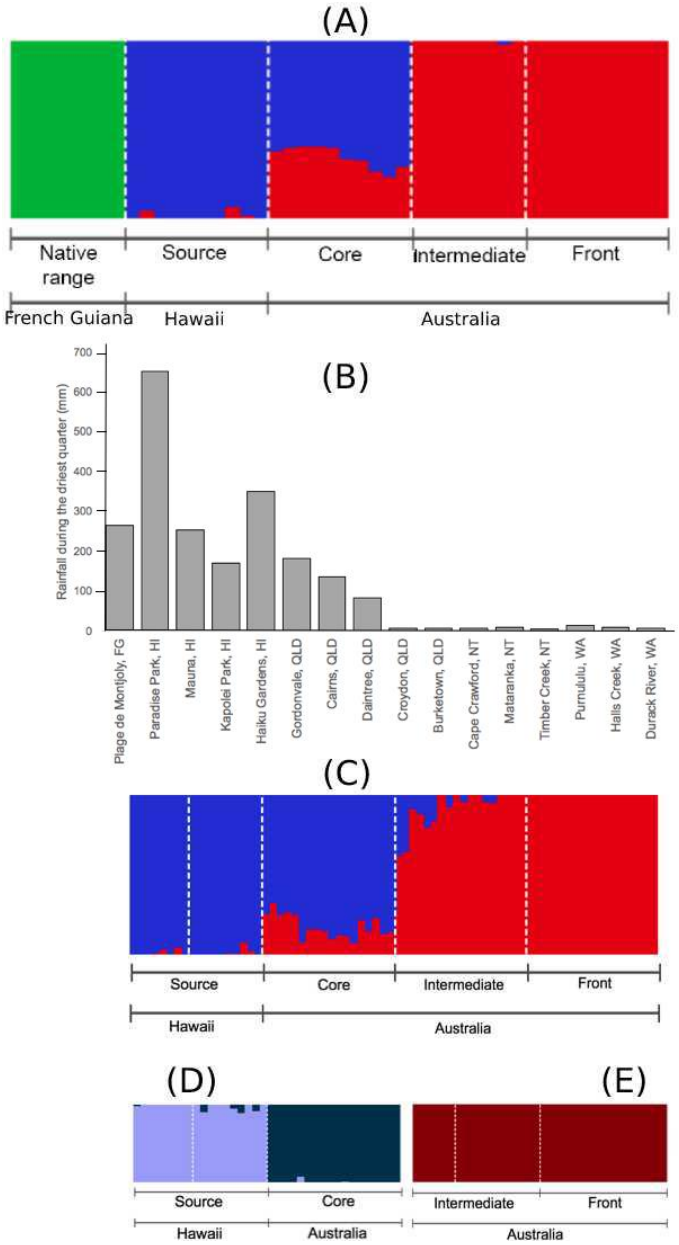

Fig 2. (A) Population structure of cane toads from French Guiana, Hawai'i, and Australia, as inferred by SNPs from spleen RNA-Seq data. (B) Rainfall during the driest quarter in each location across collection sites for invasive cane toads in French Guiana, Hawaii, and Australia. A similar environmental barrier occurs in temperature (Fig S1). (C) Population structure of invasive cane toads from Hawai'i and Australia, as inferred by SNPs from brain RNA-Seq data. (D) Substructure within toads from Hawai'i (HI) and Queensland (QLD) as inferred by SNPs from brain RNA-Seq data. (E) Substructure within toads from Northern Territory (NT) and Western Australia (WA) as inferred by SNPs from brain RNA-Seq data.

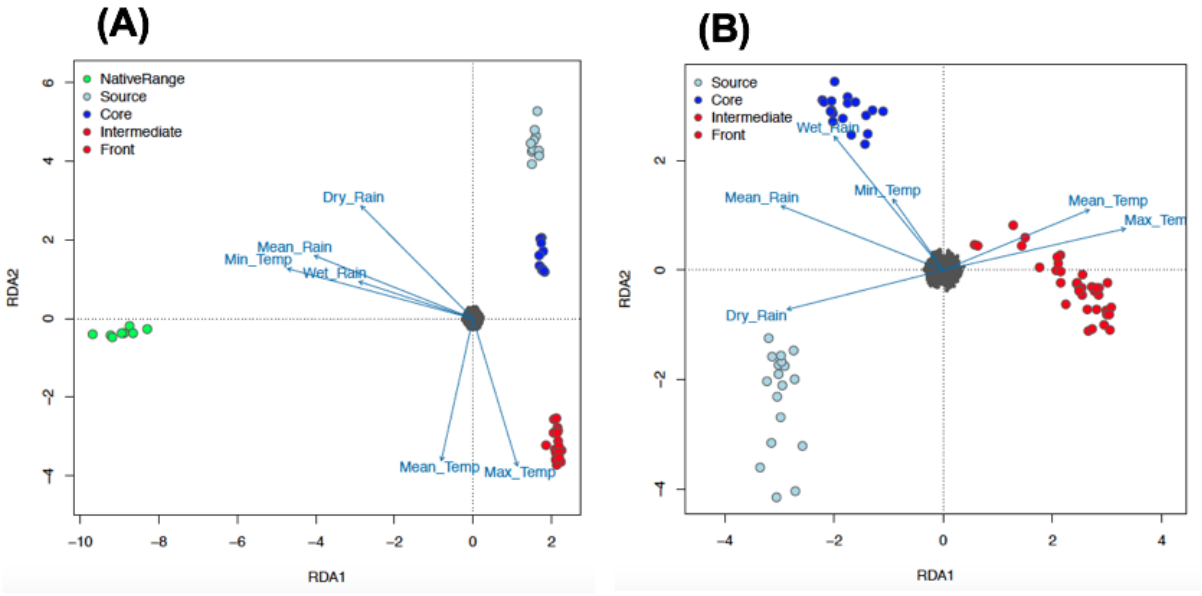

Fig 3. (A) Relationship between environmental variables and population structure of cane toads from French Guiana, Hawai'i, and Australia, as inferred by Redundancy Analysis (RDA) on SNPs from spleen RNA-Seq data. Environmental data include: annual mean temperature (Mean Temp), maximum temperature of the warmest month (Max Temp), minimum temperature of the coldest month (Min Temp), annual precipitation (Mean Rain), precipitation of the wettest quarter (Wet Rain), and precipitation of the driest quarter (Dry Rain). (B) Relationship between environmental variables and population structure of invasive cane toads from Hawai'i and Australia, as inferred by RDA on SNPs from brain RNA-Seq data.

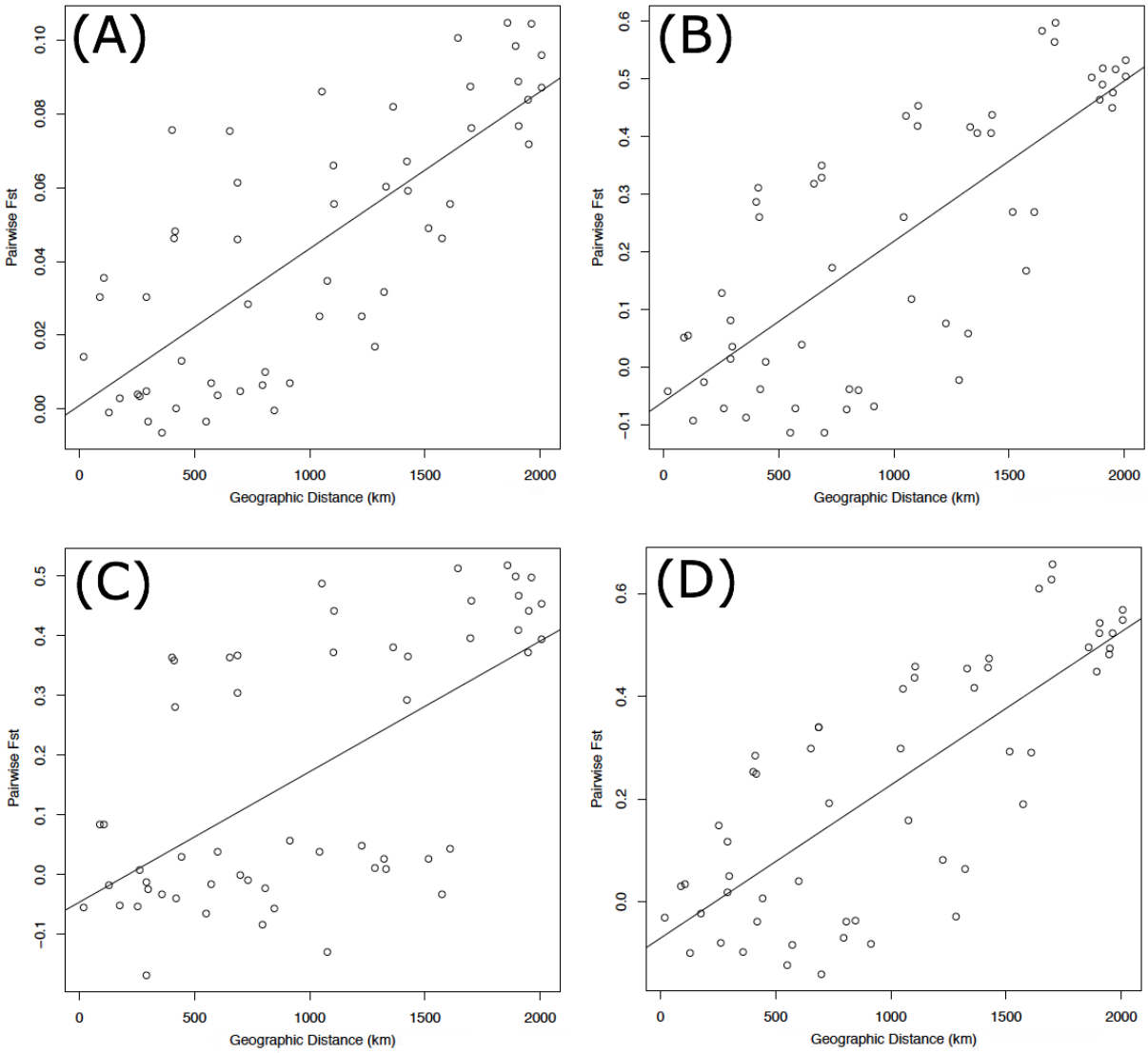

Fig 4. Relationship between geographic distance (km) and genetic distance (pairwise F_{ST}) between cane toads collected across their invasive Australian range (Fig 1) as inferred by brain RNA-Seq data. (A) represents the full dataset of loci (p = 1E-3, r = 0.76). (B) represents all F_{ST} outliers (p = 1E-3, r = 0.75). (C) represents only F_{ST} outliers with an association with either rainfall or temperature (p = 1E-3, r = 0.62). (D) represents only F_{ST} outliers without an association with either rainfall or temperature (p = 1E-3, r = 0.77).

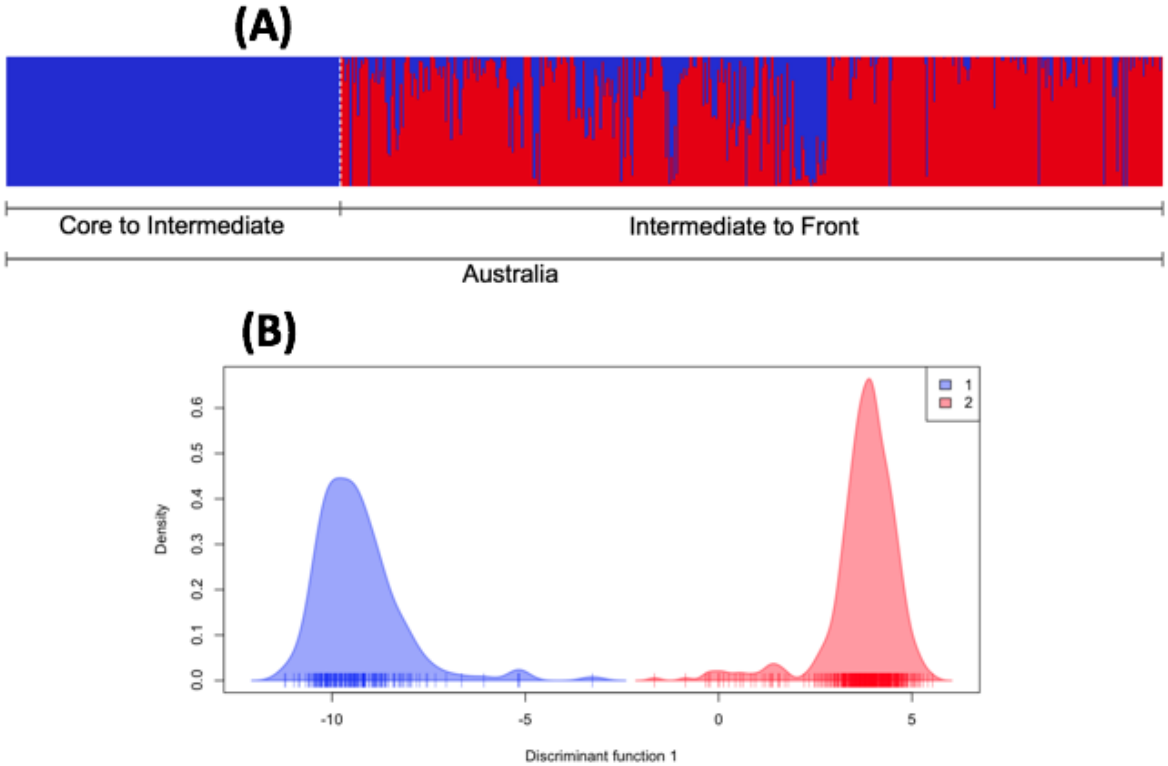

Fig 5. (A) Population structure of cane toads across their invasive range in Australia, as inferred from running fastStructure on SNPs from RADSeq data. Because coordinate metadata for collection sites were unavailable, the samples could not be arranged from east to west along the Australian range. (B) Population structure of cane toads across their invasive range in Australia, as inferred from DAPC on SNPs from RADSeq data.