FROM THE COVER

Genomic variation at the tips of the adaptive radiation of Darwin's finches

JAIME A. CHAVES,*† ELIZABETH A. COOPER,*‡ ANDREW P. HENDRY,§ JEFFREY PODOS,¶
LUIS F. DE LEÓN,**†† JOOST A. M. RAEYMAEKERS,‡‡§§ W. OWEN MACMILLAN¶¶ and
J. ALBERT C. UY*

*Department of Biology, University of Miami, Coral Gables, FL 33146, USA, †Universidad San Francisco de Quito, USFQ, Colegio de Ciencias Biológicas y Ambientales, y Extensión Galápagos, Campus Cumbayá, Quito, Ecuador, ‡Department of Genetics and Biochemistry, Clemson University, Clemson, SC 29634, USA, §Redpath Museum, Department of Biology, McGill University, Montréal, QC, Canada, ¶Department of Biology, University of Massachusetts Amherst, Amherst, MA 01003, USA, **Centro de Biodiversidad y Descubrimiento de Drogas, Instituto de Investigaciones Científicas y Servicios de Alta Tecnología (INDICASAT-AIP), Ciudad del Saber, Panama, Panama, ††Department of Biology, University of Massachusetts Boston, 100 Morrissey Blvd, Boston, MA 02125, USA, ‡‡Laboratory of Biodiversity and Evolutionary Genomics, University of Leuven, B-3000 Leuven, Belgium, §§Center for Biodiversity Dynamics, Department of Biology, Norwegian University of Science and Technology, N-7491 Trondheim, Norway, ¶¶Smithsonian Tropical Research Institute, Panama, Panama*

Abstract

Adaptive radiation unfolds as selection acts on the genetic variation underlying functional traits. The nature of this variation can be revealed by studying the tips of an ongoing adaptive radiation. We studied genomic variation at the tips of the Darwin's finch radiation; specifically focusing on polymorphism within, and variation among, three sympatric species of the genus *Geospiza*. Using restriction site-associated DNA (RAD-seq), we characterized 32 569 single-nucleotide polymorphisms (SNPs), from which 11 outlier SNPs for beak and body size were uncovered by a genomewide association study (GWAS). Principal component analysis revealed that these 11 SNPs formed four statistically linked groups. Stepwise regression then revealed that the first PC score, which included 6 of the 11 top SNPs, explained over 80% of the variation in beak size, suggesting that selection on these traits influences multiple correlated loci. The two SNPs most strongly associated with beak size were near genes associated with beak morphology across deeper branches of the radiation: *delta-like 1 homologue (DLK1)* and *high-mobility group AT-hook 2 (HMGA2)*. Our results suggest that (i) key adaptive traits are associated with a small fraction of the genome (11 of 32 569 SNPs), (ii) SNPs linked to the candidate genes are dispersed throughout the genome (on several chromosomes), and (iii) micro- and macro-evolutionary variation (roots and tips of the radiation) involve some shared and some unique genomic regions.

Keywords: adaptive radiation, beak size, Darwin's finches, genomic regions, RAD-seq

Received 23 November 2015; revision received 27 June 2016; accepted 28 June 2016

Introduction

Adaptive radiation – the diversification of a single ancestral species into multiple descendent species adapted to

Correspondence: Jaime A. Chaves, Fax: (+593 2) 289 0070;
E-mail: jachaves@usfq.edu.ec
[Correction added on 06 August, 2016, after first online publication: In the author by-line, the following author's names "JOHN. ALBERT CRUZ. UY and WILLIAM. OWEN MACMILLAN" are modified as J. ALBERT C. UY and W. OWEN MACMILLAN]

different environments – unfolds as selection acts on the genetic variation underlying adaptive traits (Huxley 1942; Futuyma 1986; Schluter 2000). It is becoming increasingly clear that the efficacy and outcome of this process depend heavily on a lineage's genetic and genomic properties. For instance, the genomic architecture (i.e. numbers, effect sizes and distributions of genes) underlying the expression of adaptive traits can constrain or facilitate their response to selection (Schluter 2000; Nosil 2012) and, reciprocally, the action of selection

can shape this architecture (Sucena *et al.* 2003; Manceau *et al.* 2010; Arnegard *et al.* 2014; Soria-Carrasco *et al.* 2014). These interactions remain poorly understood (Nosil 2012), yet important insights are emerging through advances in next-generation sequencing for non-model organisms (Kronforst *et al.* 2006; Nosil *et al.* 2007, 2009; Via & West 2008; Feder *et al.* 2011; Nosil & Feder 2012; Parchman *et al.* 2013; Lamichhaney *et al.* 2015). With these advances, we can begin to combine data on the genetic variation that underlies adaptive phenotypes with information on the intensity and nature of selection acting on those traits (Hoekstra & Nachman 2003; Hoekstra *et al.* 2006; Heliconius Genome Consortium, 2012; Nadeau *et al.* 2013; Supple *et al.* 2013).

If we are to understand the *process* of adaptive radiation, we must study its continuing action, as opposed to only examining well-established lineages that speciated in the distant past (Coyne & Orr 2004; Via 2009; Nosil & Feder 2012). To do so, we can focus on young species found at the most recent tips of the phylogeny; especially those occurring in sympatry, where polymorphisms can be most effectively studied within and among species. With this focus, we can identify genomic regions, and potential candidate genes, targeted by natural selection during adaptive divergence – without the complication of long periods of postspeciation divergence (Coyne & Orr 2004; Via 2009; Nosil & Feder 2012). The genomic architecture revealed by analyses at these recent tips then can be compared to genomic architecture across deeper branches of the radiation. Comparing these two scales facilitates inferences about how the genomic architecture underlying micro-evolutionary differences (within and among populations or very young species) relates to genetic differences that characterize larger macro-evolutionary differences (across major branches of a clade). In particular, we can assess whether the same genomic regions are associated with divergence at the two scales, providing insights into whether or not the same variation remains available to shape evolution through the course of a radiation.

The adaptive radiation of Darwin's finches is ongoing (Grant 1981; Grant & Grant 2008a). The group is composed of 14 recognized species that diverged from a common ancestor following colonization of the Galapagos approximately 1.6–2 million years ago (Petren *et al.* 2005). The different species are currently specialized for different food resources and have correspondingly appropriate beak size and shapes (Grant 1999; Grant & Grant 2008a). In a large-scale comparative study of the entire radiation, several genomic regions were implicated in beak shape variation across deep branches (i.e. across species and islands), including regions near genes that mediate craniofacial or beak development (e.g. *DLK1*, *ALX1*: Lamichhaney S, *et al.*

(2015)). More recently, variation in *HMGA2* was implicated in shaping beak and body size variation across shallower branches of the radiation (i.e. between closely related species in sympatry) and in experiencing natural selection in a population of medium ground finches *Geospiza fortis* (Lamichhaney *et al.* 2016). We here complement this work through analyses of genomic variation associated with beak and body size in sympatric species at the most recent tips of this ongoing adaptive radiation. In particular, distinct and alternative morphs within the medium ground finch on Santa Cruz Island are maintained by ongoing diversifying selection and assortative mating (Hendry *et al.* 2009) in the face of periodic bouts of hybridization and introgression (Grant 1993; Grant *et al.* 2004, 2005; Grant & Grant 2008a,b; De Léon *et al.* 2010).

Our three focal sympatric *Geospiza* ground finch species differ in beak and body size reflecting adaptation to feeding on different seed sizes. The species thus form a morphological continuum in which the small *G. fuliginosa* transitions into the medium *G. fortis*, which then transitions into the large *G. magnirostris*. Adjacent species along this continuum hybridize (Grant 1999; De Léon *et al.* 2010; Farrington *et al.* 2014), but to a limited extent owing to assortative mating based on beak size (Podos 2001; Grant & Grant 2009). At an even finer scale along this continuum, Santa Cruz *G. fortis* are further divided into small and large beak-size morphs that appear to represent incipient divergence in sympatry. The two morphs feed on seeds of different size (De Léon *et al.* 2010), mate assortatively (Huber *et al.* 2007) and show differences in feeding performance (Herrel *et al.* 2005a,b, 2009) that generate disruptive selection on beak size (Hendry *et al.* 2009; De León *et al.* 2014). Because selection on beak traits is the key mechanism driving evolution of the entire group, analysing this sympatric continuum should inform the genetic architecture associated with the process of (ongoing) adaptive radiation.

We analysed this within/among species continuum through a genomewide association study (GWAS) (Hirschhorn & Daly 2005; Parchman *et al.* 2012; Comeault *et al.* 2014) that controls for population structure. Identifying SNPs associated with variation in beak and body size, we address long-standing questions regarding the genomic architecture of adaptive divergence: How much of the variation in phenotype can be explained? How many SNPs are associated with divergence and where in the genome are they found? Do the SNPs appear to be associated with candidate genes of importance in the radiation of Darwin's finches? These results generate insights into the genomic architecture associated with trait variation at the tips of an adaptive radiation – and the association between micro- and macro-evolutionary divergence.

Methods

Genetic material and sequencing of the RAD Tag Library

We used samples collected from live birds captured in mist nets in 2011 at El Garrapatero, Santa Cruz, as part of an ongoing long-term research project (Herrel *et al.* 2005a,b, 2009; Huber *et al.* 2007; Hendry *et al.* 2009; De León *et al.* 2010; Podos 2010; J. A. M. Raeymaekers, L. F. De León, J. A. Chaves, D. M.T. Sharpe, S. K. Huber, A. Herrel, B. Vanhooydonck, J. A. H. Koop, S. A. Knutie, D.H. Clayton, R. B. Grant, P. R. Grant, J. Podos & A. P. Hendry in preparation). Each bird was classified to species based on beak and body size, the standard and effective method for Darwin's finches (Lack 1947; Grant 1999; Foster *et al.* 2008). For sequencing, we selected individuals of the small (*G. fuliginosa*; $n = 20$), medium (*G. fortis*; $n = 55$) and large (*G. magnirostris*; $n = 12$) ground finches so as to span the entire range of beak size. The larger sample size within *G. fortis* allowed us to select individuals representing the entire morphological range, including large, medium and small beak-size morphs of this highly variable species (Fig. 1).

Blood samples collected in the field were preserved in Longmire's buffer (Longmire *et al.* 1997), and genomic DNA was later extracted using commercially available extraction kits (Qiagen, Valencia, CA). Four multiplexed RAD tag libraries were constructed following a modified version of the protocol described by Parchman *et al.* (2013) (see Appendix S1, Supporting information). Sequencing was conducted at the Hussman Institute of Human Genetics, University of Miami, on an Illumina HiSeq 2000, which produced 100 bp paired-end reads. Raw Illumina reads were sorted by barcode and filtered for quality (any read with a phred quality < 20 was removed) using process_radtags from the STACKS package (Catchen *et al.* 2011). Filtered RAD tags were aligned to the *Geospiza fortis* reference genome (Zhang *et al.* 2012) using Burrows–Wheeler alignment tool (BWA) (Li & Durbin 2009), with up to 5 mismatches allowed.

SNPs were called used the *vcfutils* program in the SAMTOOLS package (Foster *et al.* 2008), with a command line option to call variants using Bayesian inference and to report likelihoods. This was followed by a filtering option to remove genotypes with a quality score less than 20. Potentially false heterozygous calls were further filtered using an in-house script to remove sites with less than 5× coverage and to replace any heterozygous calls with fewer than 20% alternate alleles with the reference allele. The slightly low coverage cut-off of 5× was selected by examining the empirical distribution of coverage in our aligned data (Fig. S6, Supporting information), where it was observed that the majority of our sites had relatively low coverage. Based on this distribution, we also selected 50× as the maximum read depth. We removed any sites with more than two alleles, less than 5% minor allele frequency, or fewer than 10 individuals per population (83% of individuals in the smallest population). These procedures were similar
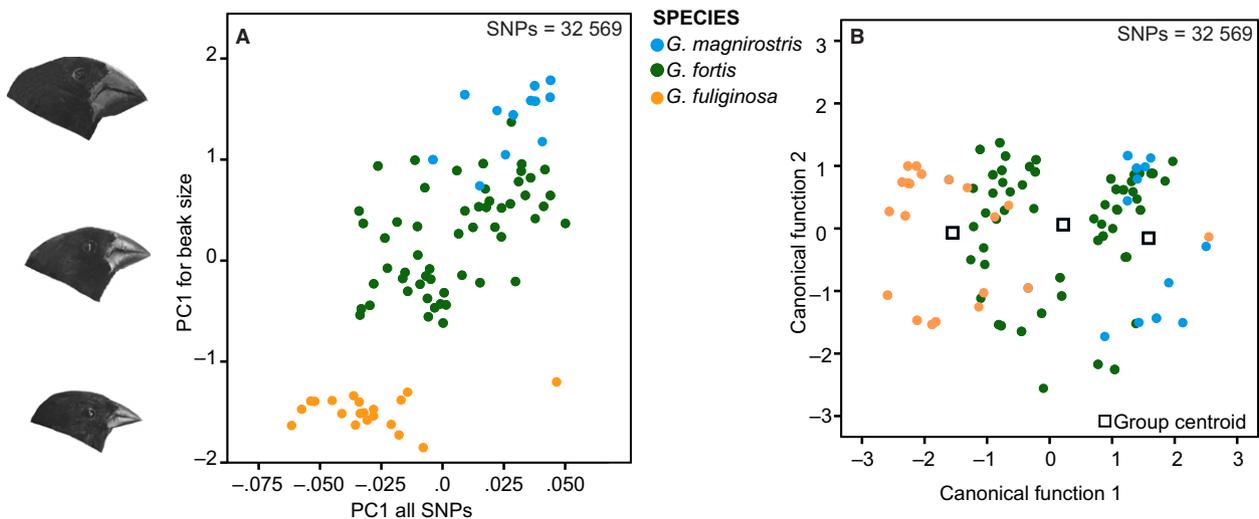
Fig. 1 Plots for principal component analysis (PCA) and discriminant analysis of principal components (DAPC) of the total panel of SNPs for 87 individuals. (A) PC1 of all 32 560 SNPs for beak size ($PC1_{ALL\_SNPs}$ = 83.29% of variance explained) on PC1 of morphology ($PC1_{BEAK\_SIZE}$ = 88.23% variance explained) ($r = 0.716$; $P = 0.000$. (B) DAPC and group centroids for the three species of Darwin's finches.

to the methods and cut-offs used by Rheindt, Cuervo and Brumfield (Rheindt *et al.* 2013), but we were more conservative because we had lower overall coverage in our data. More details about our pipeline and the scripts can be found online at: https://github.com/eacooper400/RAD_Pipeline_Info.

Morphological data and analysis

Measures of body size included tarsus length (nearest 0.1 mm with calipers), wing length (nearest 0.5 mm with calipers) and body mass (nearest 0.1 g with digital scale). Measures of beak morphology included length (nostrils to tip of beak), depth (perpendicular from nostril to lower mandible) and width (base of lower mandible). Each beak measurement was taken three times to the nearest 0.1 mm using calipers, and the median measurement was used for analyses. From these measurements, we calculated a single index of body size using a PCA on the body measurements and a single index of beak size using PCA on the beak measurements. These two principal components each explained over 90% of the variance and the two PC1s (body and beak) were highly correlated (Pearson's correlation coefficient: 0.975; $P < 0.001$). Given the equivalency of these two metrics, we henceforth used the beak size PC1 for the all analyses (Also we did not analyse beak shape PC2 as it varies little in the study species (Foster *et al.* 2008). These analyses were performed using SPSS v.21 (IBM Corp).

Population structure and genetic assignment

Estimates of F_{ST} between species pairs (*G. fortis*, *G. magnirostris*, *G. fuliginosa*) were calculated on a site-by-site basis as $F_{ST} = (H_T - H_S)/H_T$; where H_T and H_S were based on transformations of the MAF (minor allele frequencies). Because F_{ST} can be influenced by differing levels of variation both within and between species that occur by chance across the genome, we also estimated D_{XY} as the average number of pairwise differences between sequences in the different species (Cruickshank & Hahn 2014). Pairwise nucleotide divergence (D_{XY}) between *G. fuliginosa* and *G. magnirostris* was calculated using a custom R script, which used the following equation to estimate D_{XY} based on polarized minor allele frequencies (i.e. the frequency is always in reference to the same allele in both species):

$$D_{XY} = \frac{(p * (1 - q)) + (q * (1 - p))}{n}$$

where p is the minor allele frequency in *G. fuliginosa*, and q is the minor allele frequency in *G. magnirostris*. The numerator of the equation was summed over all SNP sites within a given window and then divided by

n, which is the length of the sequence (which we defined as the number of base pairs in a window with at least $30\times$ mapped sequence coverage). This calculation was performed in nonoverlapping windows of 50 kb.

We used a PCA as performed in the PLINK software (Purcell *et al.* 2007) to characterize genetic variation based on the full 32 569 SNPs data set. This analysis produced 2 PC scores, where $PC1_{ALL\_SNPs}$ explained 83.29% of the allelic variance across species. We then tested the ability of the corresponding PCs scores to effectively classify individuals into their own species using a discriminant analysis of principal components (DAPC).

Genomewide associations analysis and beak-size prediction

We scanned for SNPs associated with PC1 of beak size using the Bayesian sparse linear mixed model (BSLMM) (Zhou & Stephens 2012, 2014) implemented in the software package GEMMA (Zhou *et al.* 2013). This genomewide association approach detects SNPs that are associated with a given trait, while explicitly controlling for population structure by incorporating a relatedness matrix as a covariate in the mixed model. For each trait ($PC1_{BODY\_SIZE}$ and $PC1_{BEAK\_SIZE}$), we performed 10 independent runs of the model, each with 5 million burn-in runs followed by 20 million iterations. The hyperparameters were averaged across runs, and the mean posterior inclusion probability (PIP) values were plotted for every SNP (~30 k) across the different scaffolds. As a conservative approach to identify SNPs that were significantly associated with beak and body size, we filtered for candidate SNPs using a very strict PIP > 0.1 [X. Zhou, personal communication (Comeault *et al.* 2014)]. This cut-off is an order of magnitude higher than the widely used PIP > 0.01 (see Comeault *et al.* 2014; Gompert *et al.* 2013) and reduces the likelihood of uncovering spurious associations between SNPs and phenotype (i.e. type II error).

Because SNPs that are fixed or nearly fixed in the extreme species may be likely to be heterozygous in the intermediate species simply under Hardy–Weinberg equilibrium, and not because they actually have an associated and additive effect with beak size, we performed randomization tests to see how often our top two SNPs (*DLK1* and *HMGA2*) were significantly associated with beak size within the intermediate *G. fortis* species. The methods and results of these tests are detailed in the Supporting information. Briefly, we found that our observed associations were marginally significant for both genes, even within *G. fortis* alone.

As the SNPs uncovered by the GWAS could represent statistically linked groups, we identified

independent indices that could represent statistically correlated groups of SNPs using a PCA on the top 11 SNPs for beak size (Zhen & Altman 2004; Parchman *et al.* 2013). We then used forward stepwise regression analysis to explore the contribution of the PC_{TOP_SNPs} scores explaining beak-size variation. These analyses were performed using SPSS v.21 (IBM Corp).

SNPs associated with beak size and linkage disequilibrium

Candidate genes linked to the 11 candidate SNPs were identified by scanning the most currently available annotation of the *Geospiza* genome (Zhang *et al.* 2012) as chromosomal locations on each scaffold. In order to determine the appropriate window size to identify genes near our top SNPs, we examined linkage disequilibrium (*LD*) decay profiles. *LD* between markers was calculated as the correlation coefficient (r^2) between pairs of genotypes using the 'genetics' package in R (Warnes *et al.* 2013), which uses maximum-likelihood estimation for uncertain haplotypes. For each scaffold containing a SNP in the top 0.1% of the BSLMM iterations (see genomewide association analysis and beak-size prediction), *LD* was estimated between every possible pair of bi-allelic sites. Distances between pairs of sites were binned in increments of 100 bp, and a mean r^2 value was calculated for each bin in order to plot and evaluate the decay of *LD* with distance. While the pattern differed slightly between scaffolds, it appeared that *LD* typically began to decay (r^2 became less than 0.1) after approximately 20–40 kb, yet remained slightly elevated ($r^2 > 0.05$) even after several hundred kb (Fig. S4, Supporting information). Since genes under selection also could have elevated *LD*, we considered 200-kb windows when searching for potential candidate genes adjacent to our top SNPs. To estimate linkage between the top SNPs uncovered by the GWAS, haplotypes were first inferred using the software package PHASE (Stephens *et al.* 2001; Stephens & Donnelly 2003), with 20 independent starts of the EM algorithm. The *LD* estimators *D'* and r^2 were then calculated directly from the inferred haplotypes using a custom R script. Significance was assessed using Fisher's exact test as implemented in R.

Morphology and allelic variants across the root of the radiation

To compare the specific allelic variants across the entire radiation, we first sequenced individuals of the cactus finch (*Geospiza scandens*: *n* = 8), the sister species from Santa Cruz, following the same RAD Tag Library procedure presented above. To complete the taxonomic sampling, we aligned data files (BAM format) for additional island populations of each *G. fortis*, *G. fuliginosa*, *G. magnirostris* and *G. scandens*, as well as 2 outgroup populations of *L. noctis* and *T. bicolor* from the mainland. These data were obtained from Lamichhaney *et al.* (2015). The downloaded BAM files were merged with samtools, and SNPs were called with the program FREE-BAYES (Garrison & Marth 2012) with parameters similar to those used in our original SNP calling pipeline. Briefly, these parameters included a minimum alternate allele fraction of 20% (-F 0.2), a minimum mapping quality of 50 (-m 50), a minimum base quality of 30 (-q 30) and a minimum coverage of 20. These new SNP calls were merged with our original RAD-seq SNPs using the vcf-merge function in the VCFTOOLS package (Danecek *et al.* 2011).

Results and discussion

Our core analyses focused on 87 individuals spanning the range of beak size across the three *Gesopiza* ground finch species found in sympatry at a single location (El Garrapatero on Santa Cruz Island). Using RAD-seq (Baird *et al.* 2008; Hohenlohe *et al.* 2010, 2011), we characterized 32 569 SNPs distributed across the genome. On average, these SNPs showed little differentiation among the species, with the vast majority of markers showing $F_{ST} < 0.05$ and $D_{XY} < 0.0016$ (Figs S1 and S2, Supporting information). Differentiation was especially low for pairwise species comparisons that involved *G. fortis*, which is intermediate in beak and body size to the other two species. In short, these species are definitely found at the tips of the adaptive radiation, and so the genomic architecture of their adaptive traits should reflect the variation actively shaping – and being shaped by – natural selection.

Despite this very low average divergence, some genomic differences were present among the species. In a principal component analysis (PCA), most of the allelic variation among species was explained by $PC1_{ALL\_SNPs}$ (83.29%), which was closely correlated with beak size ($PC1_{BEAK\_SIZE}$; $r = 0.716$; $P = 0.000$; Fig. 1). This association suggests a certain degree of population structure resulting from nonrandom mating between birds that differ in beak morphology. Discriminant analysis of principal components (DAPC) confirmed the continuous nature of this beak size-associated population structure, with high assignment success for species at the two ends of the beak-size continuum (*G. magnirostris* 100%, *G. fuliginosa* 85%) but low success for the intermediate species (*G. fortis* = 32.7%). These results follow logically from the fact that *G. fortis* is highly variable and intermediate in beak size and should therefore be more likely to hybridize with the two more extreme

species than those two species are to hybridize with each other (De Léon *et al.* 2010; Farrington *et al.* 2014; Lamichhaney *et al.* 2015). As a result, *G. fortis* is expected to carry a mixture of the alleles found in the smaller and larger species. Importantly, however, no fixed differences were observed between any of the species (max $F_{ST} < 0.9$; Fig. 2, bottom panel), confirming that divergence is ongoing (or, in any case, incomplete).

Our primary goal was to identify genomic regions associated with beak size while controlling for the above-mentioned population structure. To this end, we used Bayesian sparse linear mixed model (BSLMM), a hybrid between a linear mixed model that incorporates a relatedness matrix to control for population structure and a sparse regression designed to better detect polygenic effects (rather than just single site associations) (Price *et al.* 2010; Zhang *et al.* 2010; Zhou *et al.* 2013). This analysis revealed that even our relatively sparse set of SNPs explained more than 90% of the variation in beak size ($PC1_{GWAS/BEAK}$: 94.7% [89.9–97.9] variance explained) and body size ($PC1_{GWAS/BODY}$: 93% [88.4–97.6] variance explained) (Table S1, Supporting information). Thus, 'missing heritability' and variants with undetectably small effects (see Rockman 2012) were not a concern in our study. The very high PVE values observed in our analyses probably reflect not only the

very high heritability of these morphological traits (e.g. beak size $h^2 = 0.97$ (in Boag & Grant PR (1978)), but also the likely action of several genes of very large effect (Peichel *et al.* 2001; Griswold 2006). Indeed, a small handful of SNPs showed very strong associations with trait variation. Using a posterior inclusion probability (PIP) greater then 0.1, 15 SNPs were associated with beak and body size, of which eight were associated only with beak size and four only with body size (Fig. 2 top panel, Fig. S4 and Tables S4 and S5, Supporting information). Thus, three SNPs were associated with both traits, which is expected given the very high phenotypic correlation between beak and body size ($n = 87$ birds, $r = 0.975$, $P < 0.001$; Fig. S3, Supporting information). We focus the remainder of our analyses and discussion on beak-associated SNPs only: that is, eight that predicted beak size alone and three that predicted both beak and body size.

We first determined the extent to which these 11 SNPs were independent (e.g. not in linkage disequilibrium, *LD*) by calculating pairwise *LD* (Table S2, Supporting information) and performing a PCA (PCA_{TOP_SNPs}) to derive independent axes of genetic variation (as in 16, 56). This PCA extracted four axes that together explained 66.1% of the variation in genotype, suggesting that the top 11 SNPs associated with
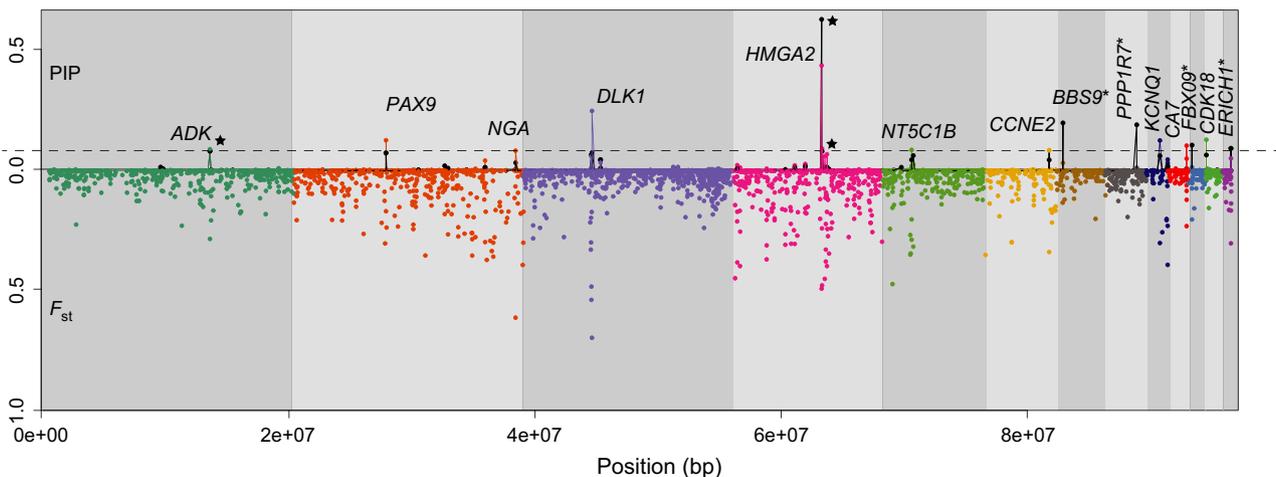
Fig. 2 Genomewide association scan for SNPs predicting beak and body size of 87 individuals of ground finches. Posterior inclusion probability (PIP) values from a Bayesian sparse linear mixed model analysis (BSLMM) for 32 569 SNPs mapped to the *Geospiza fortis* reference genome are shown on top. The horizontal dashed line represents the cut-off value of PIP = 0.1 (positive values y-axis – higher panel). Genomewide F_{ST} values of SNP from species comparisons between *G. fuliginosa*, *G. fortis* and *G. magnirostris* (positive values y–axis – lower panel). Genomic positions are indicated on the x-axis, with different scaffolds indicated by different colours and vertical grey shading. Note that the current assembly of *G. fortis* has scaffolds sorted by size, and not necessarily their putative mapping locations. Therefore, for points shown in the same colour there is genomic data suggesting that they are physically linked. In contrast, there is no evidence that points shown in different colours are physically near each other, even if they are displayed adjacent to each other in this figure. Only scaffolds containing at least one SNP with a PIP > 0.1 are shown. Second scaffold contains more than one group of linked SNPs (groups 2 and 3) and the positions of the 15 significant SNPs – above dashed line – correspond to annotated SNPs as in Table S4 and S5, Supporting information (only physically closer SNP to actual candidate gene name given). Coloured dots above the line correspond to beak-size SNPs only, whereas black dots correspond to body size only (names with asterisks), and stars correspond to overlapping beak and body size SNPs. (NGA = no gene associated identified).

beak size represented at least four independent groups of statistically linked SNPs (Table S3, Supporting information). Within each of the four groups, the SNPs were highly correlated (r^2 and D' significantly deviate from zero, $P < 0.001$; Table S2, Supporting information) despite not being found physically close to one another, which suggests a role for correlated selection (Lande & Arnold 1983). Using forward stepwise regression, the variance in beak size explained by these four SNP groups was in cumulative order 56.8% (PC1$_{TOP\_SNPs}$), 79.9% (PC1 and PC2$_{TOP\_SNPs}$), 82.3% (PC1, PC2 & PC3$_{TOP\_SNPs}$) and 83.6% (All 4 PC$_{TOP\_SNPs}$; Table 1). Removing the three overlapping SNPs (beak and body SNPs) strongly diminished the amount of variation explained (10% of the variation in genotype), highlighting the importance of those three SNPs in driving both traits.

Given that PC1$_{TOP\_SNPs}$ alone explained 56.8% of the variance in beak size, we next focused on the six SNPs that loaded significantly on this axis (see Table 1; Fig. 3). These SNPs mapped to separate scaffolds of the *G. fortis* genome (Zhang *et al.* 2012) but were in *LD* (Table S2, Supporting information): that is, they are found on different chromosomes but are statistically associated. Notably, the top SNP (PIP > 0.1) aligned to a region inside of the *high-mobility group AT-hook 2 (HMGA2)* on chromosome 1 (Table S4, Supporting information). Suggestively, the *HMGA2* protein is linked to adipogenesis and acts as an architectural transcriptional factor (Markowski *et al.* 2011) influencing body size, including skeletal development in humans (Weedon *et al.* 2007, 2008; Soranzo *et al.* 2009), mice (Zhou *et al.* 1995) and chickens (Song *et al.* 2011). Also, *HMGA2* was recently linked to variation in beak size across the entire Darwin's finch clade (Lamichhaney *et al.* 2015). Not surprisingly, then, our candidate SNP inside the intron of this gene also ranked first in a GWAS for body size (Table S5, Supporting information). Additionally, *BMP4*, previously identified as a

'beak gene' in Darwin's finch cranial development (Abzhanov *et al.* 2004), influences the activation and proliferation of *HMGA2* (Markowski *et al.* 2011). Thus, our top SNP supports assertions of the importance of *HMGA2* in beak development across the entire Darwin's finch radiation (Lamichhaney *et al.* 2016).

The second most significant SNP was 170-kb upstream of the *delta-like 1 homologue (DLK1)* on chromosome 5, which is a member of the epidermal growth factor (*EGF*)-like protein family and is expressed in a variety of tissues during vertebrate embryonic development (Shin *et al.* 2008; Falix *et al.* 2013). Suggestively, the majority of the transcription regulatory element of *DLK1* is localized *ca.* 40-kb upstream of the gene (Rogers *et al.* 2012). This SNP also showed the highest F_{ST} between *G. fuliginosa* and *G. magnirostris* in our entire data set ($F_{ST} = 0.86$) (Fig. S1, Supporting information). Although the SNP was also associated with body size ($0.01 < PIP < 0.1$), it appears to be more specifically associated with beak size (PIP = 0.29; Table S4, Supporting information). Suggestively, *DLK1* has recently been linked to variation in beak shape across the entire Darwin's finch clade, and it is strongly associated with other cranial regulatory factors (Lamichhaney *et al.* 2015).

Although none of our SNPs are probably *causal*, they are strongly associated with beak size presumably due to physical proximity to candidate genes. We therefore explored in greater detail the association between morphology and specific allelic variants for the above two SNPs. For both SNPs, most *G. fuliginosa* were homozygous and had the same alleles (T at the SNP associated with *HMGA2* and G at the SNP associated with *DLK1*), with only one bird in each group homozygous for the alternative *DLK1* allele and one bird homozygous for the alternate *HMGA2* allele (Table 2). In sharp contrast, nearly all sequenced *G. magnirostris* were homozygous for a different allele at both SNPs in the alternative state (C at the SNP associated with *HMGA2* SNP linked to

Table 1 Stepwise forward regression analysis and performance of PC$_{TOP-SNPs}$ axes predicting beak size

Model	R	R square	Adjusted R square	Std. error of the estimate	Change statistics				
					R square change	F change	Df1	Df2	Sig. F change
1	0.757*	0.573	0.568	0.6609	0.573	114.152	1	85	0.000*
2	0.896[†]	0.803	0.799	0.4514	0.230	98.207	1	84	0.000[†]
3	0.911[‡]	0.830	0.823	0.4226	0.026	12.854	1	83	0.001[‡]
4	0.919[§]	0.844	0.836	0.4071	0.014	7.460	1	82	0.008[d]

*Predictors: (constant), PC1.
[†]Predictors: (constant), PC1 and PC2.
[‡]Predictors: (constant), PC1, PC2 and PC3.
[§]Predictors: (constant), PC1, PC2, PC3 and PC4.
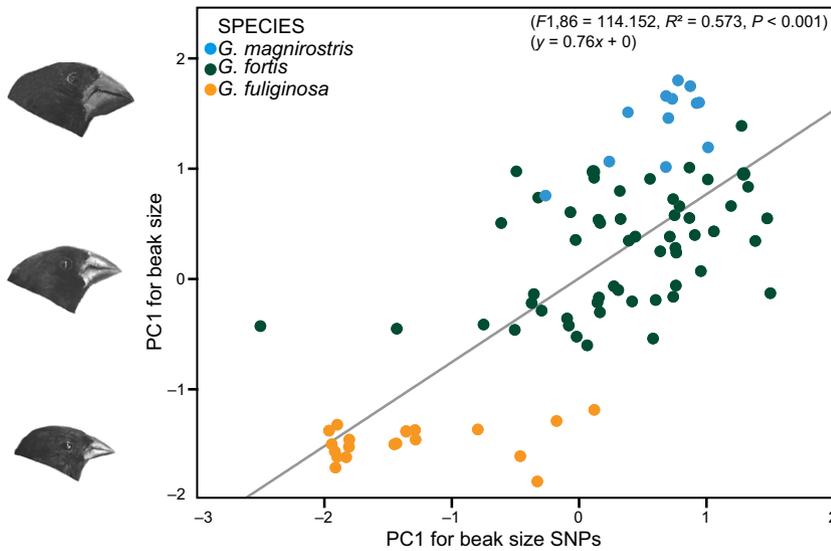
Fig. 3 Regression of $PC1_{BEAK\_SIZE}$ (88.23% variance explained) on PC1 for beak-size SNPs ($PC1_{TOP\_SNPs}$ = 56.8% variance explained) of the top candidate SNPs.

HMGA2 and A at the SNP associated with *DLK1* SNP linked to *DLK1*), with only one bird heterozygous for *HMGA2*. Between these two extremes, *G. fortis* birds had a variety of genotypes – and these were associated with beak size. That is, *G. fortis* individuals homozygous for the same alleles as most *G. fuliginosa* were significantly smaller than *G. fortis* individuals homozygous for the same alleles as most *G. magnirostris* alternate alleles; moreover, heterozygous *G. fortis* were intermediate in beak size (Table 2, Fig. 4A and B). Notably, some genotype combinations were never observed in our data or in previously published data: for example, no individuals homozygous for the 'small' *HMGA2*-associated allele ever possessed a 'large' alternative *DLK1*-associated allele. This strong *LD* despite physical independence of the two SNPs (which are not themselves causal) suggests they might be reliable markers for inferring correlational selection on particular allelic combinations at the nearby candidate loci.

We first plotted median trait values as a function of each pairwise genotype (Fig. S5, Supporting information) and found that each locus appears to be mostly, although not entirely, associated with additive effects – most obviously for body size. For the exceptions, slight recessiveness was suggested for the 'large' allele at the *DLK1*-associated SNP and the 'small' allele at the *HMGA2*-associated SNP. The trends with beak size were a bit more complicated and may be indicative of some epistatic interactions, but could also be confounded by our small sample size. A type III ANOVA (to account for unequal group sizes) returned no significant interaction between the two loci (P = 0.89), but this test may not be entirely reliable given that some groups, or genotype pairs, are actually completely missing in our data set.

We therefore further explored potential interactions between alleles at the two SNPs. All individuals that were homozygous for the 'small' allele at the *HMGA2*-associated SNP were also homozygous for the 'small' allele at the *DLK1*-associated SNP and, when heterozygous, a given allele ('small' vs. 'large') at the one SNP was never found with the alternative allele ('large' vs. 'small') at the other SNP. This cross-locus allelic association also held true in publicly available data collected by other groups (see below), further supporting the above suggestions of correlational selection maintaining *LD* across genomic regions. However, it is also possible that alleles at the two SNPs are maintained in *LD* by independent (as opposed to correlational) selection at each of the associated candidate genes. That is, divergent selection shaping 'small–small' and 'large–large' allelic combinations across the two SNPs (owing to selection at the nearby candidate genes) might simply be the sum of selection at the two independent loci. Simulations – and studies of natural selection on polymorphic populations in nature – could help to distinguish these possibilities.

To explore in more detail allelic variation – and its potential origins – at the SNPs near *HMGA2* and *DLK1*, we genotyped cactus finches (*Geospiza scandens*) at El which is sister to the three sympatric ground finch species of Santa Cruz Island (Petren *et al.* 1999). We also examined published data on the genotypes of SNPs associated with *HMGA2* and *DLK1* from Darwin's finches from other islands and from outgroups (*Tiaris noctis* and *Tiaris bicolor*, both of which have small beaks) to *Geospiza* found on the mainland (Lamichhaney *et al.* 2015). These data (Table 2) support our findings. First, *G. magnirostris* birds from Genovesa Island were homozygous for the same 'large' variant at both SNPs.

Table 2 Allele variants for candidate genes associated with out two top SNPs (*HMGA2* and *DLK1*) across the radiation. Species sampled correspond both to the tips of the radiation (*G. fuliginosa*, *G. fortis* (included 3 morphs) and *G. magnirostris*) from Santa Cruz (this study) and to the root of the radiation from sister taxa across other islands, included the outgroups (*: from Lamichhaney *et al.* 2015). Letters correspond to either homozygous (same) or heterozygous (different) individuals, and number of counts, for individuals sampled (*n*)

	HMGA2	*DLK1*
G. fuliginosa (Santa Cruz)	TT (*n* = 14)	GG (*n* = 17)
G. fuliginosa (Santa Cruz)	CT (*n* = 5)	GA (*n* = 1)
G. fuliginosa (Santa Cruz)	CC (*n* = 1)	AA (*n* = 1)
G. fortis (small morph) (Santa Cruz)	TT (*n* = 2)	GG (*n* = 19)
G. fortis (intermediate morph) (Santa Cruz)	CT (*n* = 17)	GA (*n* = 9)
G. fortis (large morph) (Santa Cruz)	CC (*n* = 36)	AA (*n* = 24)
G. magnirostris (Santa Cruz)	CC (*n* = 11)	AA (*n* = 12)
G. magnirostris (Santa Cruz)	CT (*n* = 1)	GA (*n* = 0)
G. magnirostris (Santa Cruz)	TT (*n* = 0)	GG (*n* = 0)
G. scandens (Santa Cruz)	TT (*n* = 3)	GG (*n* = 7)
G. scandens (Santa Cruz)	CT (*n* = 5)	GA (*n* = 1)
G. scandens (Santa Cruz)	CC (*n* = 0)	AA (*n* = 0)
G.magnirostris (Genovesa)*	CC (*n* = 1)	AA (*n* = 1)
G. fortis (Daphne)*	CT (*n* = 1)	GA (*n* = 1)
G. scandens (Daphne)*	TT (*n* = 2)	GG (*n* = 2)
G. fuliginosa (Santiago)*	TT (*n* = 2)	GG (*n* = 2)
G. fuliginosa (Santa Cruz)*	TT (*n* = 2)	GG (*n* = 2)
Tiaris noctis (mainland)*	TT (*n* = 5)	GG (*n* = 5)
T. bicolor (mainland)*	TT (*n* = 3)	GG (*n* = 3)

Second, *G. scandens* at El Garrapatero and all *G. fuliginosa*, *G. fortis* (with one exception on Daphne Major) and *G. scandens* from other islands, as well as the outgroups, were homozygous for the same variants as *G. fuliginosa* and small *G. fortis* from Santa Cruz. Thus, alleles T at the *HMGA2*-associated SNP and G at the *DLK1*-associated SNP could be the ancestral for the entire radiation. Under this scenario, C at the *HMGA2*-associated SNP and A at the *DLK1*-associated SNP found could have evolved in *G. magnirostris* on Galapagos and facilitated (through hybridization) the evolution of large *G. fortis* birds on Santa Cruz (Table 2, Fig. 4). This suggestion fits with observations that hybridization is not uncommon between *G. fortis* and *G. magnirostris* (Grant 1999; Grant & Grant 2008a). Furthermore, our discriminant analysis suggests unidirectional gene flow from *G. magnirostris* to *G. fortis* as the former (but not the latter) possesses 100% assignment to its own species group (probably due to private alleles).

Moving beyond the above two 'top' SNPs, several other SNPs from the first PC_{TOP_SNPs} axis, as well as some SNPs from the other three PC_{TOP_SNPs} axes, also showed intriguing associations. For instance, one of the

SNPs that loaded onto $PC1_{TOP\_SNPs}$ also mapped near *HMGA2* and is in close physical proximity (*ca.* 30 kb) to the above-analysed SNP (Table S4, Supporting information). Additionally, a number of SNPs loading on the other axes map to regions of the genome near loci that regulate foetal development, skeletal elements formation and body size (Tables S2, S4 and S5, Supporting information). Finally, it is valuable to reiterate that three of our top 11 SNPs predicted both beak and body size, which is expected given the very high phenotypic correlation between beak and body size (i.e. allometry), and so, we speculate that some of the same genes probably influence multiple correlated aspects of morphology.

Comparison to independent studies

Our analysis of variation at the tips of an adaptive radiation, including distinct beak polymorphism within a single species, provides interesting comparisons to independent findings across the entire Darwin's finch clade (Lamichhaney *et al.* 2015, 2016). For instance, a comparative genomics study of *G. conirostris*, *G. difficilis*, *G. magnirostris* and *G. fortis* implicated *ALX1* in beak shape divergence (Lamichhaney *et al.* 2015), as well as other craniofacial regulator genes such as *Goosecoid homeobox* (*GSC*) (Rivera-Perez *et al.* 1999), which is presumably very near or in high *LD* with *DLK1* (Lamichhaney *et al.* 2015). Further, different levels of *BMP4* and *CaM* expression have been shown to influence beak size and shape in *G. fuliginosa*, *G. fortis*, *G. magnirostris* and the cactus finch *G. scandens* (Abzhanov *et al.* 2004, 2006). By contrast, none of our top 11 candidate SNPs were found near *ALX1*, *GSC*, *BMP4* and *CaM*. Our results therefore suggest that some genes are shared between the roots and tips of a radiation (as described above), whereas others differ between those two scales.

One possible explanation for apparent discrepancies between the different studies could be the limited genomic coverage of our RAD-seq approach, as well as other methodological differences (although our parallel results for *HMGA2*-associated SNPs (see Lamichhaney S, *et al.* (2016)) indicate that our different approaches can yield very similar outcomes). First, our marker density may have been too low to provide reliable markers for the above candidate genes. Arguing against this possibility, we had SNPs located *ca.* 2-kb downstream of *BMP4*, *ca.* 800-bp upstream of *ALX1* and *ca.* 33-kb upstream of *GSC*. Moreover, our genomewide measurements of *LD* decay across all scaffolds were *ca.* 20–40 kb (Fig. S4, Supporting information) confirming the lack of association between our physically close SNPs to previously defined 'beak genes' and beak-size differences observed on Santa Cruz. Second, the focus of our study was on beak size, whereas the other studies mostly focused on beak shape. In our
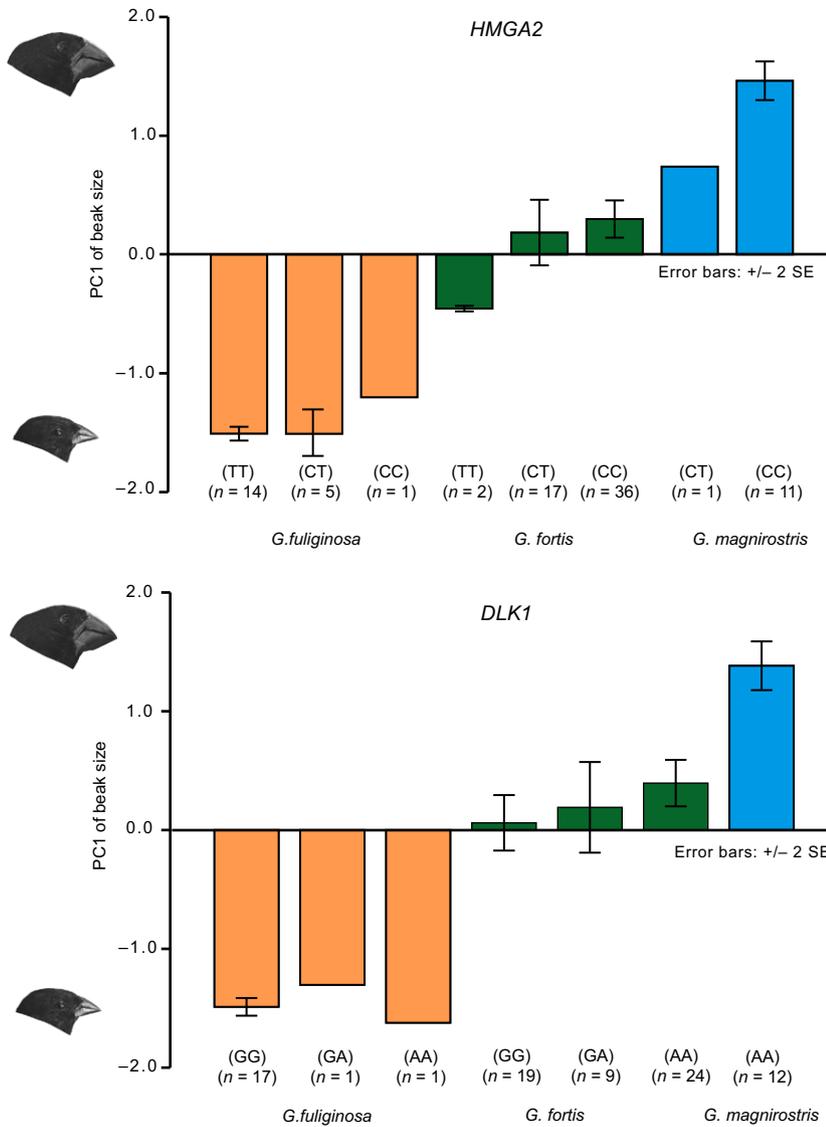
Fig. 4 Beak-size association with candidate SNPs in *Geospiza* finches. Plot of PC1 of beak size and its corresponding allelic variants for SNPs associated with *HMGA2* and *DLK1*.

sample, 83.29% of the variation in beak dimensions was size, leaving little residual variation in beak shape (see also Campás *et al*. 2010; Foster *et al*. 2008). At the same time, however, candidate genes mediating size and shape might not always be distinct, with a possible case being *HMGA2* and *DLK1*.

In Lamichhaney *et al*. (2015, 2016), *HMGA2*, *DLK1* and *GSC* were well differentiated across the entire Darwin's finch clade, and, as noted above, two of these genes are physically close to our two most important SNPs. Thus, these genes might be associated with both current microevolutionary variation (polymorphism at the tips of the branches) and past macro-evolutionary variation (deeper splits among species and genera) in the same adaptive radiation. This finding highlights the importance of standing variation in the process of intra- and interspecific diversification and also suggests that hybridization between species, as previously shown between *G. fortis*

and *G. magnirostris* on Santa Cruz, may be important for maintaining this variation and allowing continued adaptation. Overall, then, micro- and macrolevels of differentiation probably involve a mixture of shared and unique genes, depending on how populations diverge.

Conclusion

Based on a panel of over 30 000 SNPs, our results suggest that only handful of SNPs predicted the majority of variation in key phenotypic traits (83.6%, PC_{TOP_SNPs}), after controlling for population structure. As such, one of the key results of our study is that a surprisingly modest proportion of SNPs (i.e. 0.03% of total SNPs) mediate beak size across the beak-size gradient in sympatry, most likely due some genes of large effect as previously reported in this radiation (Abzhanov *et al*. 2004, 2006; Lamichhaney *et al*. 2015, 2016). Further, we found that

the predictive SNPs were found spread throughout the genome (on several chromosomes) but were in *LD*, presumably due to selection. Finally, the few strongly differentiated SNPs were located in or near a number of candidate genes previously implicated in Darwin's finches adaptive radiation. We therefore suggest that some, but not all, of the genomic regions involved in the deep branches of a radiation (macroevolution) are involved in ongoing divergence at the tips of the radiation (microevolution). This could result in genes involved in the early phases of adaptive radiation to remain available for later phases of adaptive radiation (i.e. they are not constrained). The consequences for trait evolution in this group imply that variation in these genes is maintained either as standing variation or through new mutations or through hybridization. Introgressive hybridization could then be responsible for increasing genetic variation providing opportunity for disruptive selection to mediate the early stages of ecological speciation at the tips of the radiation.

Acknowledgements

Conflict of interests

The authors declared that they have no conflict of interests.

References

Abzhanov A, Protas M, Grant BR, Grant PR, Tabin CJ (2004) Bmp4 and morphological variation of Beaks in Darwin's finches. *Science*, **305**, 1462–1465.

Abzhanov A, Kuo WP, Hartmann C, Grant RB, Grant PR, Tabin CJ (2006) The calmodulin pathway and evolution of elongated beak morphology in Darwin's finches. *Nature*, **442**, 563–567.

Arnegard ME, McGee MD, Matthews B *et al.* (2014) Genetics of ecological divergence during speciation. *Nature*, **511**, 307–311.

Baird NA, Etter PD, Atwood TS *et al.* (2008) Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PLoS ONE* **3**, e3376.

Boag PT, Grant PR (1978) Heritability of external morphology in Darwin's finches. *Nature*, **274**, 793–794.

Campàs O, Mallarino R, Herrel A, Abzhanov A, Brenner MP (2010) Scaling and shear transformations capture beak shape variation in Darwin's finches. *Proceedings of the National Academy of Sciences*, **107**, 3356–3360.

Catchen JM, Amores A, Hohenlohe P, Cresko W, Postlethwait JH (2011) Stacks: building and genotyping loci de novo from short-read sequences. *G3: Genes Genomes Genetics*, **1**, 171–182.

Comeault AA, Soria-Carrasco V, Gompert Z *et al.* (2014) Genome-wide association mapping of phenotypic traits subject to a range of intensities of natural selection in timema cristinae. *The American Naturalist*, **183**, 711–727.

Coyne JA, Orr HA (2004) *Speciation*. Sinauer Associates Inc., Sunderland, Massachusetts.

Cruickshank TE, Hahn MW (2014) Reanalysis suggests that genomic islands of speciation are due to reduced diversity, not reduced gene flow. *Molecular Ecology*, **23**, 3133–3157.

Danecek P, Auton A, Abecasis G *et al.* (2011) The variant call format and VCFtools. *Bioinformatics*, **27**, 2156–2158.

De León LF, Podos J, Gardezi T, Herrel A, Hendry AP (2014) Darwin's finches and their diet niches: the sympatric coexistence of imperfect generalists. *Journal of Evolutionary Biology*, **27**, 1093–1104.

De Léon LF, Bermingham E, Podos J, Hendry AP (2010) Divergence with gene flow as facilitated by ecological differences: within-island variation in Darwin's finches. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **365**, 1041–1052.

Falix FA, Tjon-A-Loi MRS, Gaemers IC, Aronson DC, Lamers WH (2013) DLK1 protein expression during mouse development provides new insights into its function. *ISRN Developmental Biology*, **2013**, 10.

Farrington HL, Lawson LP, Clark CM, Petren K (2014) The evolutionary history of Darwin's Finches: speciation, gene flow, and introgression in a fragmented landscape. *Evolution*, **68**, 2932–2944.

Feder JL, Egan SP, Nosil P (2011) The genomics of speciation-with-gene-flow. *Trends in Genetics*, **28**, 342–350.

Foster DJ, Podos J, Hendry AP (2008) A geometric morphometric appraisal of beak shape in Darwin's finches. *Journal of Evolutionary Biology*, **21**, 263–275.

Futuyma DJ (1986) *Evolutionary Biology*, 2nd edn. Sinauer, Mass

Garrison E, Marth G (2012) Haplotype-based variant detection from short-read sequencing. arXiv:1207.3907 [q-bio.GN].

Gompert Z, Lucas LK, Nice CC, Buerkle CA (2013) Genome divergence and the genetic architecture of barriers to gene flow between Lycaeides and *L. melissa*. *Evolution*, **67**, 2498–2514.

Grant PR (1981) Speciation and the adaptive radiation of Darwin's finches. *American Scientist*, **60**, 653–663.

Grant PR (1993) Hybridization of Darwin's Finches on Isla Daphne Major, Galapagos. *Philosophical Transactions of the*

Royal Society of London. Series B: Biological Sciences, **340**, 127–139.

Grant PR (1999) *Ecology and Evolution of Darwin's Finches*, 2nd edn. Princeton University Press, Princeton, New Jersey.

Grant PR, Grant BR (2008a) *How and why Species Multiply: The Radiation of Darwin's Finches*. Princeton University Press, Princeton, New Jersey.

Grant BR, Grant PR (2008b) Fission and fusion of Darwin's Finches populations. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **363**, 2821–2829.

Grant PR, Grant BR (2009) Sympatric speciation, immigration, and hybridization in island birds. In: *The Theory of Island Biogeography Revisited* (eds Losos JB, Ricklefs RE), pp 326–357, Princeton University Press, Princeton, New Jersey.

Grant PR, Grant BR, Markert JA, Keller LF, Petren K (2004) Convergent evolution of Darwin's Finches caused by introgressive hybridization and selection. *Evolution*, **58**, 1588–1599.

Grant PR, Grant BR, Petren K (2005) Hybridization in the recent past. *The American Naturalist*, **166**, 56–67.

Griswold CK (2006) Gene flow's effect on the genetic architecture of a local adaptation and its consequences for QTL analyses. *Heredity*, **96**, 445–453.

Heliconius Genome Consortium (2012) Butterfly genome reveals promiscuous exchange of mimicry adaptations among species. *Nature*, **487**, 94–98.

Hendry AP, Huber SK, De León LF, Herrel A, Podos J (2009) Disruptive selection in a bimodal population of Darwin's finches. *Proceedings of the Royal Society B: Biological Sciences*, **276**, 753–759.

Herrel A, Podos J, Huber SK, Hendry AP (2005a) Bite performance and morphology in a population of Darwin's finches: implications for the evolution of beak shape. *Functional Ecology*, **19**, 43–48.

Herrel A, Podos J, Huber SK, Hendry AP (2005b) Evolution of bite force in Darwin's finches: a key role for head width. *Journal of Evolutionary Biology*, **18**, 669–675.

Herrel A, Podos J, Vanhooydonck B, Hendry AP (2009) Force–velocity trade-off in Darwin's finch jaw function: a biomechanical basis for ecological speciation? *Functional Ecology*, **23**, 119–125.

Hirschhorn JN, Daly MJ (2005) Genome-wide association studies for common diseases and complex traits. *Nature Reviews Genetics*, **6**, 95–108.

Hoekstra HE, Nachman MW (2003) Different genes underlie adaptive melanism in different populations of rock pocket mice. *Molecular Ecology*, **12**, 1185–1194.

Hoekstra HE, Hirschmann RJ, Bundey RA, Insel PA, Crossland JP (2006) A single amino acid mutation contributes to adaptive beach mouse color pattern. *Science*, **313**, 101–104.

Hohenlohe PA, Bassham S, Etter PD, Stiffler N, Johnson EA, Cresko WA (2010) Population genomics of parallel adaptation in threespine stickleback using sequenced RAD tags. *PLoS Genetics*, **6**, e1000862.

Hohenlohe PA, Amish SJ, Catchen JM, Allendorf FW, Luikart G (2011) Next-generation RAD sequencing identifies thousands of SNPs for assessing hybridization between rainbow and westslope cutthroat trout. *Molecular Ecology Resources*, **11**, 117–122.

Huber SK, León LFD, Hendry AP, Bermingham E, Podos J (2007) Reproductive isolation of sympatric morphs in a population of Darwin's finches. *Proceedings of the Royal Society B: Biological Sciences*, **274**, 1709–1714.

Huxley J (1942) *Evolution, the Modern Synthesis*. Allen & Unwin, London.

Kronforst MR, Young LG, Kapan DD, McNeely C, O'Neill RJ, Gilbert LE (2006) Linkage of butterfly mate preference and wing color preference cue at the genomic location of wingless. *Proceedings of the National Academy of Sciences*, **103**, 6575–6580.

Lack DL (1947) *Darwin's Finches*. Cambridge University Press, Cambridge, UK.

Lamichhaney S, Berglund J, Almén MS et al. (2015) Evolution of Darwin's finches and their beaks revealed by genome sequencing. *Nature*, **518**, 371–375.

Lamichhaney S, Han F, Berglund J et al. (2016) A beak size locus in Darwin's finches facilitated character displacement during a drought. *Science*, **352**, 470–474.

Lande R, Arnold SJ (1983) The measurement of selection on correlated characters. *Evolution*, **37**, 1210–1226.

Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*, **25**, 1754–1760.

Longmire JL, Maltbie M, Baker RJ (1997) Use of "lysis buffer" in DNA isolation and its implication for museum collections. *Occassional Papers of the Museum of Texas Tech University*, **163**, 1–3.

Manceau M, Domingues VS, Linnen CR, Rosenblum EB, Hoekstra HE (2010) Convergence in pigmentation at multiple levels: mutations, genes and function. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **365**, 2439–2450.

Markowski DN, Helmke BM, Meyer F et al. (2011) BMP4 increases expression of HMGA2 in mesenchymal stem cells. *Cytokine*, **56**, 811–816.

Nadeau NJ, Martin SH, Kozak KM et al. (2013) Genome-wide patterns of divergence and gene flow across a butterfly radiation. *Molecular Ecology*, **22**, 814–826.

Nosil P (2012) *Ecological Speciation*. Oxford University Press, New York, New York.

Nosil P, Feder JL (2012) Genomic divergence during speciation: causes and consequences. *Philosophical Transactions of the Royal Society of London Series B-Biological Sciences*, **367**, 332–342.

Nosil P, Egan SP, Funk DJ (2007) Heterogeneous genomic differentiation between walking-stick ecotypes:"Isolation by adaptation" and multiple roles for divergent selection. *Evolution*, **62**, 316–336.

Nosil P, Funk C, Ortiz-Barrientos D (2009) Divergent selection and hetergenous genomic divergence. *Molecular Ecology*, **18**, 375–402.

Parchman TL, Gompert Z, Mudge J, Schilkey FD, Benkman CW, Buerkle CA (2012) Genome-wide association genetics of an adaptive trait in lodgepole pine. *Molecular Ecology*, **21**, 2991–3005.

Parchman TL, Gompert Z, Braun MJ et al. (2013) The genomic consequences of adaptive divergence and reproductive isolation between species of manakins. *Molecular Ecology*, **22**, 3304–3317.

Peichel CL, Nereng KS, Ohgi KA et al. (2001) The genetic architecture of divergence between threespine stickleback species. *Nature*, **414**, 901–905.

Petren K, Grant BR, Grant PR (1999) A phylogeny of Darwin's finches based on microsatellite DNA length variation. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, **266**, 321–329.

Petren K, Grant PR, Grant BR, Keller LF (2005) Comparative landscape genetics and the adaptive radiation of Darwin's finches: the role of peripheral isolation. *Molecular Ecology*, **14**, 2943–2957.

Podos J (2001) Correlated evolution of morphology and vocal signal structure in Darwin's finches. *Nature*, **409**, 185–188.

Podos J (2010) Acoustic discrimination of sympatric morphs in Darwin's finches: a behavioural mechanism for assortative mating? *Philosophical Transactions of the Royal Society B: Biological Sciences*, **365**, 1031–1039.

Price AL, Zaitlen NA, Reich D, Patterson N (2010) New approaches to population stratification in genome-wide association studies. *Nature Reviews Genetics*, **11**, 459–463.

Purcell S, Neale B, Todd-Brown K *et al.* (2007) PLINK: a toolset for whole-genome association and population-based linkage analysis. *American Journal of Human Genetics*, **81**, 559–575.

Rheindt FE, Cuervo AM, Brumfield RT (2013) Rampant polyphyly indicates cryptic diversity in a clade of Neotropical flycatchers (Aves: Tyrannidae). *Biological Journal of the Linnean Society*, **108**, 889–900.

Rivera-Perez JA, Wakamiya M, Behringer RR (1999) Goosecoid acts cell autonomously in mesenchyme-derived tissues during craniofacial development. *Development*, **126**, 3811–3821.

Rockman MV (2012) The QTN program and the alleles that matter for evolution: all that's gold does not glitter. *Evolution*, **66**, 1–17.

Rogers ED, Ramalie JR, McMurray EN, Schmidt JV (2012) Localizing transcriptional regulatory elements at the mouse Dlk1 locus. *PLoS ONE*, **7**, e36483.

Schluter D (2000) *The Ecology of Adaptive Radiation* (Oxford Series in Ecology and Evolution).

Shin J, Lim S, Latshaw JD, Lee K (2008) Cloning and expression of delta-like protein 1 messenger ribonucleic acid during development of adipose and muscle tissues in chickens. *Poultry Science*, **87**, 2636–2646.

Song C, Gu X, Feng C *et al.* (2011) Evaluation of SNPs in the chicken HMGA2 gene as markers for body weight gain. *Animal Genetics*, **42**, 333–336.

Soranzo N, Rivadeneira F, Chinappen-Horsley U *et al.* (2009) Meta-analysis of genome-wide scans for human adult stature identifies novel loci and associations with measures of skeletal frame size. *PLoS Genetics*, **5**, e1000445.

Soria-Carrasco V, Gompert Z, Comeault AA *et al.* (2014) Stick insect genomes reveal natural selection's role in parallel speciation. *Science*, **344**, 738–742.

Stephens M, Donnelly P (2003) A comparison of Bayesian methods for haplotype reconstruction from population genotype data. *American Journal of Human Genetics*, **73**, 1162–1169.

Stephens M, Smith NJ, Donnelly P (2001) A new statistical method for haplotype reconstruction from population data. *American Journal of Human Genetics*, **68**, 978–989.

Sucena E, Delon I, Jones I, Payre F, Stern DL (2003) Regulatory evolution of shavenbaby/ovo underlies multiple cases of morphological parallelism. *Nature*, **424**, 935–938.

Supple MA, Hines HM, Dasmahapatra KK *et al.* (2013) Genomic architecture of adaptive color pattern divergence and convergence in Heliconius butterflies. *Genome Research*, **23**, 1248–1257.

Via S (2009) Natural selection in action during speciation. *Proceedings of the National Academy of Sciences*, **106**(Suppl 1), 9939–9946.

Via S, West J (2008) The genetic mosaic suggests a new role for hitchhiking in ecological speciation. *Molecular Ecology*, **17**, 4334–4345.

Warnes G, Gorjanc G, Leisch F, Man M (2013) genetics: Population Genetics (R package), 1.3.6.

Weedon MN, Lettre G, Freathy RM *et al.* (2007) A common variant of HMGA2 is associated with adult and childhood height in the general population. *Nature Genetics*, **39**, 1245–1250.

Weedon MN, Lango H, Lindgren CM *et al.* (2008) Genome-wide association analysis identifies 20 loci that influence adult height. *Nature Genetics*, **40**, 575–583.

Zang G, Parker P, Li B, Li H, Wang J (2012) The genome of Darwin's Finch (*Geospiza fortis*). *GigaScience*, doi.org/10.5524/100040.

Zhang Z, Ersoz E, Lai CQ *et al.* (2010) Mixed linear model approach adapted for genome-wide association studies. *Nature Genetics*, **42**, 355–360.

Zhen L, Altman RB (2004) Finding haplotype tagging SNPs by use of principal components analysis. *American Journal of Human Genetics*, **75**, 850–861.

Zhou X, Stephens M (2012) Genome-wide efficient mixed-model analysis for association studies. *Nature Genetics*, **44**, 821–824.

Zhou X, Stephens M (2014) Efficient multivariate linear mixed model algorithms for genome-wide association studies. *Natural Methods*, **11**, 407–409.

Zhou X, Benson KF, Ashar HR, Chada K (1995) Mutation responsible for the mouse pygmy phenotype in the developmentally regulated factor HMGI-C. *Nature*, **376**, 771–774.

Zhou X, Carbonetto P, Stephens M (2013) Polygenic modeling with bayesian sparse linear mixed models. *PLoS Genetics*, **9**, e1003264.

Data accessibility

Data set produced in this publication can be found at the Sequence Read Archive (SRA) http://www.ncbi.nlm.nih.gov/sra/SRP066465 Reference number SRP066465.

Supporting information

Additional supporting information may be found in the online version of this article.

Appendix S1 RAD tag library preparation, D_{XY} analysis, and randomization tests for *DLK1* and *HMGA2* SNPs.

Fig. S1 Frequency distribution of F_{ST} scores for all pairwise species comparisons based on 32,569 filtered SNPs. Means are marked by vertical solid black lines.

Fig. S2 D_{XY} calculated in non-overlapping 50 kb sliding windows across the 13 scaffolds which contained SNPs significantly associated with beak or body size.

Fig. S3 Correlation between PC Beak on PC Body from 87 Darwin's finches (*G. fuliginosa*, *G. fortis* and *G. magnirostris*) from Santa Cruz Island (Pearson $r = 0.975$; $P < 0.01$).

Fig. S4 Decay of linkage disequilibrium with distance in base pairs units for all species comparisons (top) and just for *G. fortis* (bottom).

Fig. S5 Interaction of *DLK1* and *HMGA2* and their individual effects on beak and body size.

Fig. S6 Empirical distribution of mapped read coverage in the full data set.

Table S1 Bayesian Sparse Linear Mixed Model output from GEMMA for PC1 beak size and PC1 body size.

Table S2 Linkage disequilibrium (*LD*) indices for pairwise comparisons of 11 top SNPs.

Table S3 Principal Component Analysis of 11 top SNPs (top) and Factor loading (below) for each PC extracted.

Table S4 Information about top SNPs in Darwin's finches genomic associations with beak size sorted by PIP values (Gamma; from highest to lowest in Bold).

Table S5 Information about top SNPs in Darwin's finches genomic associations with body size sorted by PIP values (Gamma; from highest to lowest in Bold).