Population Genomics of Variegated Toad-Headed Lizard *Phrynocephalus versicolor* and Its Adaptation to the Colorful Sand of the Gobi Desert

Yuanting Jin[1,*,†], Diana Aguilar-Gómez [2,*,†], Débora Y.C. Brandt[3], Tyler A. Square[4], Jiasheng Li[1], Zhengxia Liu[1], Tao Wang[5], Peter H. Sudmant[2,3], Craig T. Miller[4], and Rasmus Nielsen [2,3]

[1]College of Life Sciences, China Jiliang University, Hangzhou 310018, Zhejiang, China
[2]Center for Computational Biology, University of California Berkeley, Berkeley, CA, USA
[3]Department of Integrative Biology, University of California Berkeley, Berkeley, CA, USA
[4]Department of Molecular and Cell Biology, University of California Berkeley, Berkeley, CA, USA
[5]College of Life Sciences and Medicine, Zhejiang Sci-Tech University, Hangzhou 310018, China

[†]These authors contributed equally to this work.

*Corresponding authors: E-mails: aguilar-gomez@berkeley.edu; jinyuanting@126.com.

Accepted: 19 May 2022

Abstract

The variegated toad-headed agama, *Phrynocephalus versicolor*, lives in the arid landscape of the Chinese Gobi Desert. We analyzed populations from three different locations which vary in substrate color and altitude: Heishankou (HSK), Guazhou County (GZ), and Ejin Banner (EJN). The substrate color is either light-yellow (GZ-y), yellow (EJN-y), or black (HSK-b); the corresponding lizard population colors largely match their substrate in the degree of melanism. We assembled the *P. versicolor* genome and sequenced over 90 individuals from the three different populations. Genetic divergence between populations corresponds to their geographic distribution. We inferred the genetic relationships among these populations and used selection scans and differential expression to identify genes that show signatures of selection. *Slc2a11* and *akap12*, among other genes, are highly differentiated and may be responsible for pigment adaptation to substrate color in *P. versicolor*.

Key words: pigmentation, melanism, lizard, genetics.

Significance

Before this study, genetic resources for *Phrynocephalus versicolor* were limited to mitochondrial sequences. Here we present the de novo assembly of this lizard genome and resequencing data from 94 individuals from three different populations. We present an example of a workflow using different methods and statistics that can be applied to low-coverage sequencing projects that rely on genotype likelihoods.

Introduction

Pigmentation is a phenotypic trait often subjected to selective pressures via the perception of other organisms, for example, by a predator or a potential sexual partner. It can also have physiological functions such as thermoregulation and water loss reduction (Cuthill et al. 2017; Burgon et al. 2020). There are now multiple studies of the genetic basis of color variation in a variety of different species (Theron et al. 2001; Nachman et al. 2003; Rosenblum et al. 2004;

© The Author(s) 2022. Published by Oxford University Press on behalf of Society for Molecular Biology and Evolution.
This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License (https://creativecommons.org/licenses/by-nc/4.0/), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

Downloaded from https://academic.oup.com/gbe/article/14/7/evac076/6604964 by University of California, Berkeley/LBL user on 08 July 2022

Corso et al. 2012; Albertson et al. 2014; McLean et al. 2017; Burgon et al. 2020). In many cases, *mc1r* has been found responsible for changes in melanic and nonmelanic populations of reptiles and mammals that match their substrate (Nachman et al. 2003; Rosenblum et al. 2004; Corso et al. 2012; Cox et al.2013; Jin et al. 2020; Hasegawa et al. 2020). However, the pathways that produce chromatophores and their pigments involve several gene regulatory networks and enzymatic pathways comprising many different loci.

In reptiles, pigments are stored in four different types of chromatophores: (1) melanophores which contain melanin-producing black-brown colors, (2) iridophores containing light-reflecting guanine crystals which can help produce blue color, (3) xanthophores, and (4) erythrophores containing carotenoid and pteridine pigments that produce yellow-red coloration. They contain pterinosomes which are vesicles that produce pteridine pigments, and carotenoids must be obtained through diet. Xanthophores/erythrophores selectively absorb short wavelengths of light (violet-blue), allowing longer wavelengths (yellow-red) to pass through (Morrison et al. 1995; McLean et al. 2017; Hasegawa et al. 2020). The cell types are ordered spatially in the dermis such that xanthophores and erythrophores are in the upper layers, iridophores are in the middle, and melanophores are in the deepest layers (Bagnara et al. 1968; Hasegawa et al. 2020). Pigment cell distribution/density also plays an important role in coloration. For example, melanophore density is responsible for the lightness or darkness of the skin (Hasegawa et al. 2020).

The arid landscape of the Gobi Desert is inhabited by the variegated toad-headed agama, *Phrynocephalus versicolor*, a lizard that, as the name suggests (in Latin, English, and even Chinese), exhibits variation in its coloration. We analyzed populations from different locations in China, where the sand of the mountains is either black or yellow/light-yellow (fig. 1*A* and *B*). The corresponding populations exhibit dark and light skin color that may be an adaptation to avoid predator detection similar to the mechanism proposed for lizards at White Sands, New Mexico (Rosenblum et al. 2010). The dorsal coloration and reflectance differences between populations were shown to be statistically significant in a previous study (Tong et al. 2019). Notably, the phenotype of an individual is not plastic, as shown by substrate transplantation experiments done in these populations (Tong et al. 2019), suggesting that the different color morphs are mainly based on genetic differences rather than phenotypic plasticity. We performed genome-wide selection scans to identify potential adaptive trait loci. For this purpose, we assembled de novo and annotated the *P. versicolor* genome.

Results

Phenotypic Differences

In a previous study, we measured skin reflectance in different populations of *P. versicolor* and determined that individuals from nonmelanic populations have significantly higher reflectance values (fig. 1*C*). Our previous study also showed these differences are not plastic and thus are likely to have a genetic basis (Tong et al. 2019). This suggests that evolved skin pigmentation differences may exist between populations of *P. versicolor*. To understand the cellular basis of the pigmentation differences observed between the populations inhabiting the lightest and darkest substrates, we analyzed skin samples by transmission electron microscopy (TEM). Samples were taken from equivalent regions on the flank of two representative individuals of the nonmelanic (yellow) lizards from the Guazhou County (GZ-y) and two melanic (black) lizards from the region of Heishankou (HSK-b) (fig. 1). Melanophores, iridophores, and xanthophores were detected in the dermis of both populations, though not in equal amounts: as expected, the HSK-b animal presented relatively more melanophores compared with the GZ-y animals (fig. 1*D* and *E*). HSK-b melanophores were also detected in the epidermis, which were never observed in the GZ-y animals (fig. 1*F* and *G*). Additionally, the xanthophores from the GZ-y specimen contained numerous electron-dense bodies previously described as pterinosomes in other animals (Phang et al. 2012; Brejcha et al. 2019), sometimes appearing in clusters (arrows in E′ and E″), whereas these were never observed in samples from the HSK-b population. These TEM data from representative individuals thus suggest that there is a decrease in melanocytes and a change in xanthophore organelle composition in GZ-y relative to HSK-b. Although a more thorough sampling would be needed to provide statistical support for these observed differences, these TEM data suggest that evolved differences in pigment cell abundance and/or composition may underly coloration differences between melanic and nonmelanic *P. versicolor* populations.

Genomic Assembly

We decided to take a genetic approach to understand more about the divergence and coloration differences between *P. versicolor* populations. There were no genetic resources for this species available. We generated the first genomic assembly for this species by sequencing a female EJN-y individual. The raw output of the assembly from Supernova was 2.4 Gb, but contained duplicated scaffolds. The BUSCO completeness score (*C*) was 90%, with 68% single-copy orthologs (*S*), and 21% duplicated (*D*). After the processing described in Materials and Methods, including the removal of duplicated scaffolds, the assembly size was 2 Gb, with a completeness score of 90% (S:77.3%,

Downloaded from https://academic.oup.com/gbe/article/14/7/evac076/6604964 by University of California, Berkeley/LBL user on 08 July 2022

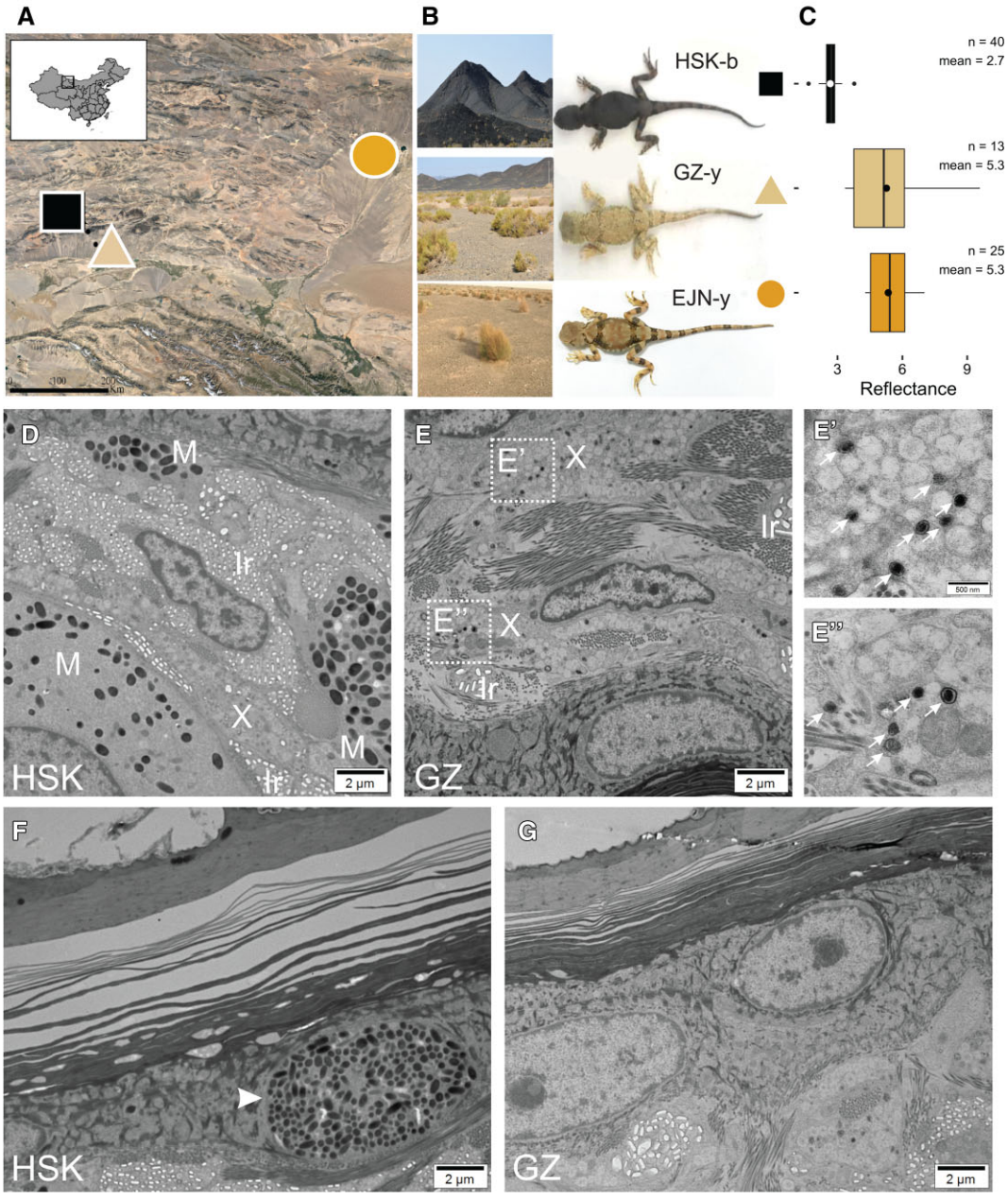

Fig. 1.—Populations of *P. versicolor* vary in melanism and exhibit different cell type and organelle composition. Individuals were collected from a melanic (HSK-b) and two nonmelanic populations (GZ-y and EJN-y). (*A*) Sampling locations are shown on the map. (*B*) Pictures of lizards and their substrate. (*C*) Reflectance measurements. (*D*) An individual from population HSK-b showed relatively more melanophores (M), whereas (*E*) an individual from population GZ-y showed relatively more xanthophores (X), which also contained some pterinosomes (E′ and E′′ show close-ups of panel E as delineated above, arrows indicate pterinosomes.) Pterinosomes were never detected in HSK-b xanthophores. Iridophores (Ir) were present in samples from both populations. (*F*) Epidermis melanophores in HSK-b. (*G*) Epidermis in GZ-y.

D:1.5%) and the N50 of 39 Mb. To enhance the computational efficacy of downstream analyses we removed scaffolds smaller than 10 kb, resulting in an assembly of 1.7 Gb. As mentioned in the Materials and Methods, we then reformatted to submit to NCBI obtaining 1.6 Gb. Although the BUSCO score improved (supplementary table S1, Supplementary Material online), the number of scaffolds was reduced from 750,174 to 4,558. The resulting assembly compares well with other published lizard assemblies (supplementary table S6, Supplementary Material online) (Alföldi et al. 2011; Georges et al. 2015; Tollis et al. 2018). We generated a draft annotation of the genome by

Downloaded from https://academic.oup.com/gbe/article/14/7/evac076/6604964 by University of California, Berkeley/LBL user on 08 July 2022

combining the transcriptome annotation and BRAKER2 results. Novogene annotated 34,185 transcripts which were aligned to the genome and used as annotation. To extract exons and corroborate genes we complemented with the BRAKER2 annotation which identified 36,433 genes.

Population Structure

To understand how populations are genetically related to each other and how divergent they are, we resequenced individuals from three populations (fig. 1*A* and *B*), one melanic population (HSK-b, $n = 44$), and two nonmelanic populations (GZ-y, $n = 22$; EJN-y, $n = 28$). We mapped them to our genome and obtained 30,657,068 single nucleotide polymorphisms (SNPs) after filtering 94 resequenced genomes of *P. versicolor*. The details of the filtering are shown in the Materials and Methods. We used this SNP set in the subsequent analyses. Using principal component analysis (PCA), population branch statistic (PBS), and the fixation index (F_{ST}), we found that the geographically closer populations GZ-y (nonmelanic) and HSK-b (melanic) are also genetically closer than either is to the nonmelanic EJN-y population (fig. 2*A* and *B*). The first principal component explains 14.7% of the variance and clearly separates EJN-y (nonmelanic) from GZ-y and HSK-b (fig. 2*B*). Individuals from different populations do not overlap on this PCA axis (fig. 2*B*). F_{ST} is higher between EJN-y and the other two populations (EJN-y-GZ-y: $F_{ST} = 0.156$, EJN-y-HSK-b: $F_{ST} = 0.175$), than it is between GZ-y and HSK-b ($F_{ST} = 0.048$). The average genomic tree estimated from the distance matrices of concatenated sequences (block size 50 Mb) is consistent with these results (supplementary fig. S6, Supplementary Material online). GZ-y and HSK-b are less divergent from each other compared with EJN-y. EJN-y and HSK-b are monophyletic in this tree. EJN-y is an outgroup to GZ-y and HSK-b and the most parsimonious hypothesis is, therefore, that the melanic skin in the HSK-b population is the derived phenotype.

In the PCA plot (fig. 2*B*) there are primarily two individuals that separate from the rest (top). Similarly, in the average genomic tree these two samples share a long internal branch in the GZ-y population but are still within the GZ-y population clades (supplementary fig. S6, Supplementary Material online). As it is only two individuals separating from the rest, we do not have sufficient data to do a more detailed analysis of the population structure. However, we did not consider that the divergence of these two individuals from the rest was sufficient grounds for exclusion.

Regions of Selection and Differential Expression Between Color Morphs

The same set of SNPs was used for scanning the genome for selection using PBS. We annotated the genomic windows using a transcriptome aligned to the genome. We looked at the top 99.5% PBS values in each population scan (fig. 2*C*, *E*, and *G*), obtaining 906 outlier transcripts for EJN-y, 634 for GZ-y, and 708 for HSK-b. The union of these transcripts resulted in 1,452 transcripts corresponding to 1,069 genes (supplementary table S7, Supplementary Material online in Excel format). Out of these 1,609 genes, 38 are differentially expressed between GZ-y and HSK-d (q-value < 0.05) according to the RNA-seq analysis (supplementary table S8 and fig. S3A, Supplementary Material online). Some of the 38 genes have been previously reported to be involved in pigmentation pathways: *gstt1* (Kanetsky et al. 2001; Polimanti et al. 2015), *slc2a11* (Kimura et al. 2014; Burgon et al. 2020), *cdkn2b* (Dorshorst and Ashwell 2009; Gryzińska et al.2013), *akap12* (Isoldi et al. 2010; Santos et al. 2016), and *scnn1b* (McLean et al. 2017; Burgon et al. 2020). One of the highest peaks in the EJN-y PBS scan contains a gene with high sequence similarity to the melanophilin gene (*mlph*) which has been previously reported in diluted color phenotypes in mammals (Ishida et al. 2006; Drögemüller et al. 2007; Demars et al. 2018). The 38 genes also included *slc2a11*, which was also an outlier in the test for reduced diversity (fig. 2*D*, *F*, and *H*). The GZ-y population has increased diversity and the EJN-y population has reduced diversity in this gene compared with other populations (fig. 2*F* and *H*). We further explore diverged paralogs of this gene (see below). We found increased Tajima's *D* for GZ-y population in this region (fig. 3*D*), which is compatible with a balancing selection hypothesis. PBS in GZ-y was not increased at this locus, consistent with the balancing selection hypothesis (fig. 2*G*). No other genes with reduced diversity displayed differential expression in the RNA-seq analyses. For example, the region containing the gene *ormdl1*, which was previously reported to affect pigmentation (Senczuk et al. 2020), has reduced diversity in the melanic HSK-b population (fig. 2*D*) but does not display differential expression. Similarly, the region containing blocs1, which is known to be involved in melanosome biogenesis (Wasmeier et al. 2008), has increased diversity in the HSK-b population (fig. 2*D*) but was not found differentially expressed.

Genomic Association

We performed a genomic association with skin reflectance to identify genetic variants associated with color. To correct for population structure, we used the loadings from the first ten PCs as covariates (fig. 3*C*). The QQplot suggests that the covariates adequately account for the latent population structure (fig. 3*B*). The SNPs with the lowest *P*-values were found in the following genes: *ninj2*, *cpas2*, *syne1*, and *pcdhgb*. We note that our data are very limited for this purpose with ~3× coverage and only 78 individuals for which

Downloaded from https://academic.oup.com/gbe/article/14/7/evac076/6604964 by University of California, Berkeley/LBL user on 08 July 2022

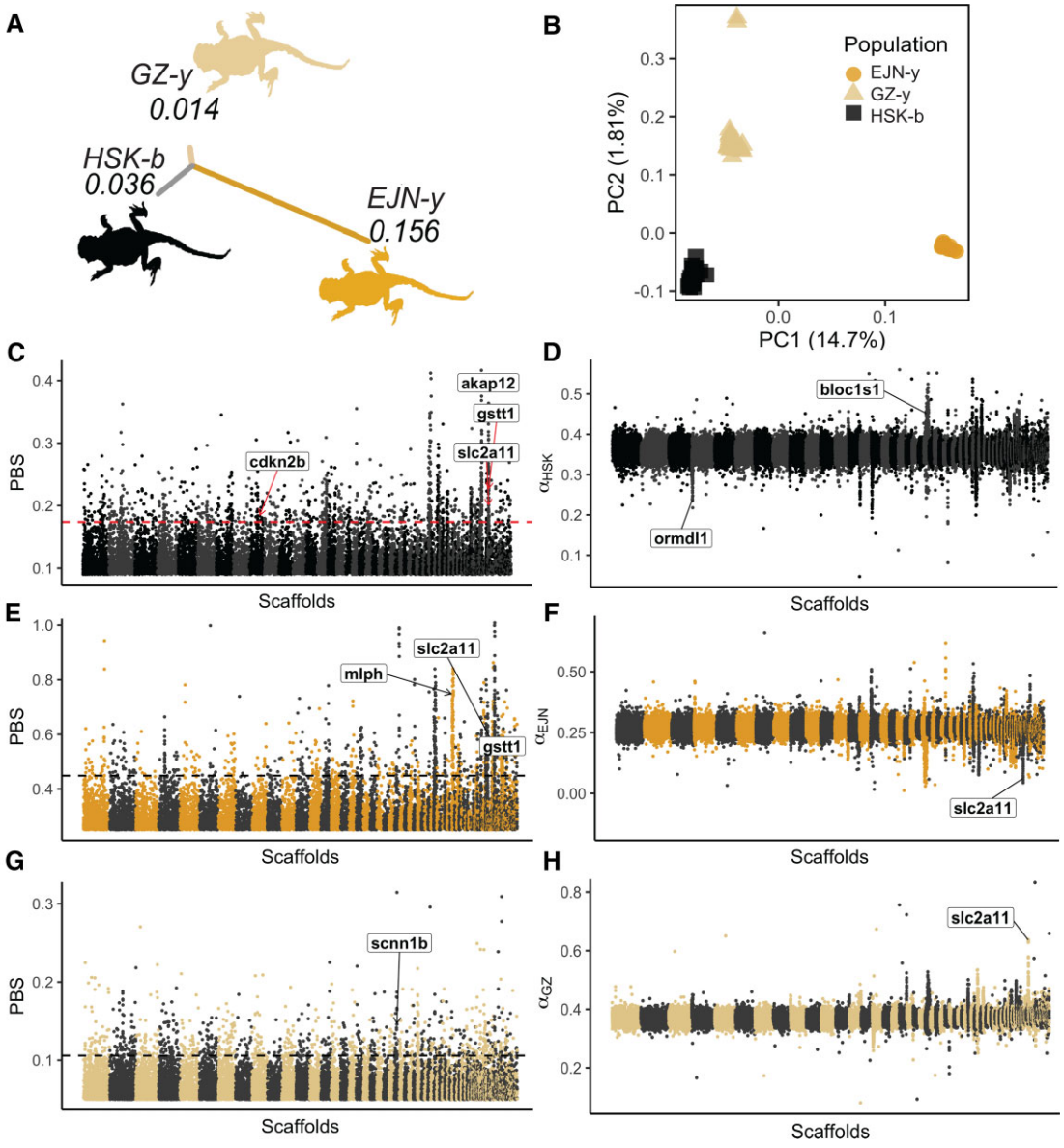

Fig. 2.—Genomic divergence and genomic scans in populations of *P. versicolor*. (*A*) Tree with genome-wide PBS branch lengths (EJN-y: 0.156, GZ-y: 0.014, HSK-b: 0.036), (*B*) PCA, (*C*) PBS scan in nonoverlapping windows for HSK-b, (*E*) PBS scan of EJN-y, (*G*) PBS scan of GZ-y. Dashed lines represent the boundary for the SNPs with the top 99.5% PBS values. (*D*, *F*, and *H*) Diversity test scan for HSK-b, EJN-y, and GZ-y, respectively. For the scans, outlier genes that are also differentially expressed, previously reported as pigmentation genes, or mentioned in the text are highlighted.

we had reflectance data and only 2,754 positions retained after hard filtering (see Materials and Methods). Overall, this genomic association is likely highly underpowered but would be able to detect genetic variants of very strong phenotypic effect, such as Mendelian phenotypes.

The highest peak of the HSK-b PBS contains the gene *akap12* (fig. 2*C* and fig. 4*A*). The lead SNPs for this locus did not pass the hard filters for coverage used in the GWAS. However, we further investigated the relationship of *akap12* to the color phenotype, by quantifying the association between genotypes in this locus and reflectance. In a genomic association performed using ten PCs as covariates, where we did not filter out sites with lower coverage, we detected five SNPs at scaffold405 inside or flanking the *akap12* gene which also had high PBS values: (1) scaffold405:492129 had a LRT value of 5.571008 which corresponds to a *P*-value of 0.01826SNP, (2) scaffold405:491094 had a LRT value of 5.580719 and *P*-value of 0.01815927, (3) scaffold405:492434 had a LRT value of 5.852483 and *P*-value of 0.01555509, (4) scaffold405:488857 had a LRT value of 11.01976 and *P*-value of 0.00090, and (5) scaffold405:489138 with LRT value

Downloaded from https://academic.oup.com/gbe/article/14/7/evac076/6604964 by University of California, Berkeley/LBL user on 08 July 2022

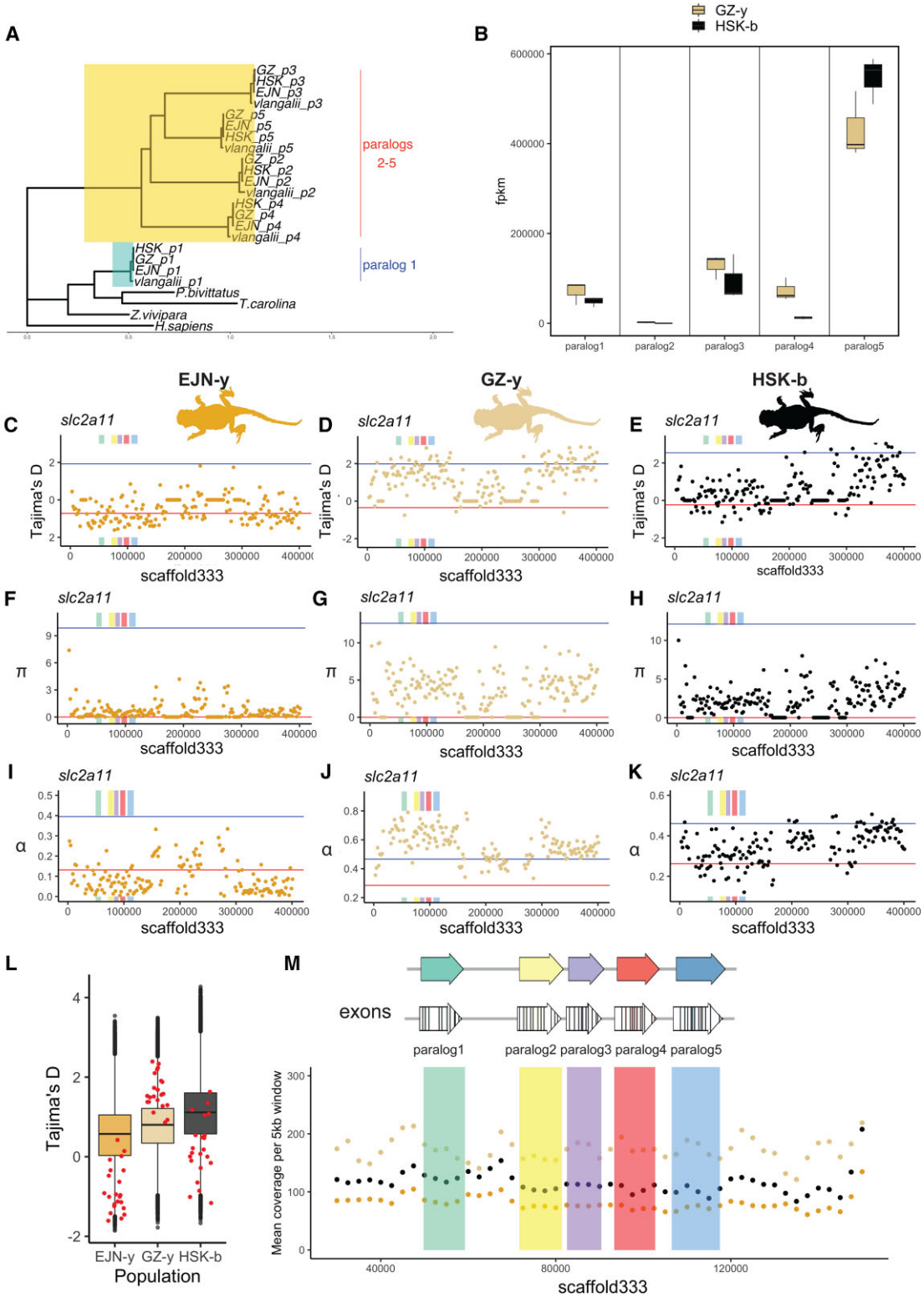

Fig. 3.—*Slc2a11* has several paralogs, is differentially expressed and shows signatures of selection. (*A*) Gene tree of slc2a11 paralogs in different *P. versicolor* morphs and in other species. (*B*) Differential expression of the different paralogs of slc2a11. (*C–K*) Zoom-in of the region containing the *slc2a11* paralogsshowing Tajima's D, pi and alpha in 2 kb windows. Columns: EJN-y, GZ-y, and HSK-bParalogs are annotated with the colors of panel M. Each plot has lines indicating the genome-wide 97.5% percentiles (blue-top, red-bottom). (*L*) Boxplot showing Tajima's *D*-values for all 2 kb windows genome wide, and 2 kb windows overlapping with any slc2a11 paralog are overlayed in red. (*M*) Diagram of the five paralogs of slc2a11 and mean coverage of populations over this locus.

Downloaded from https://academic.oup.com/gbe/article/14/7/evac076/6604964 by University of California, Berkeley/LBL user on 08 July 2022

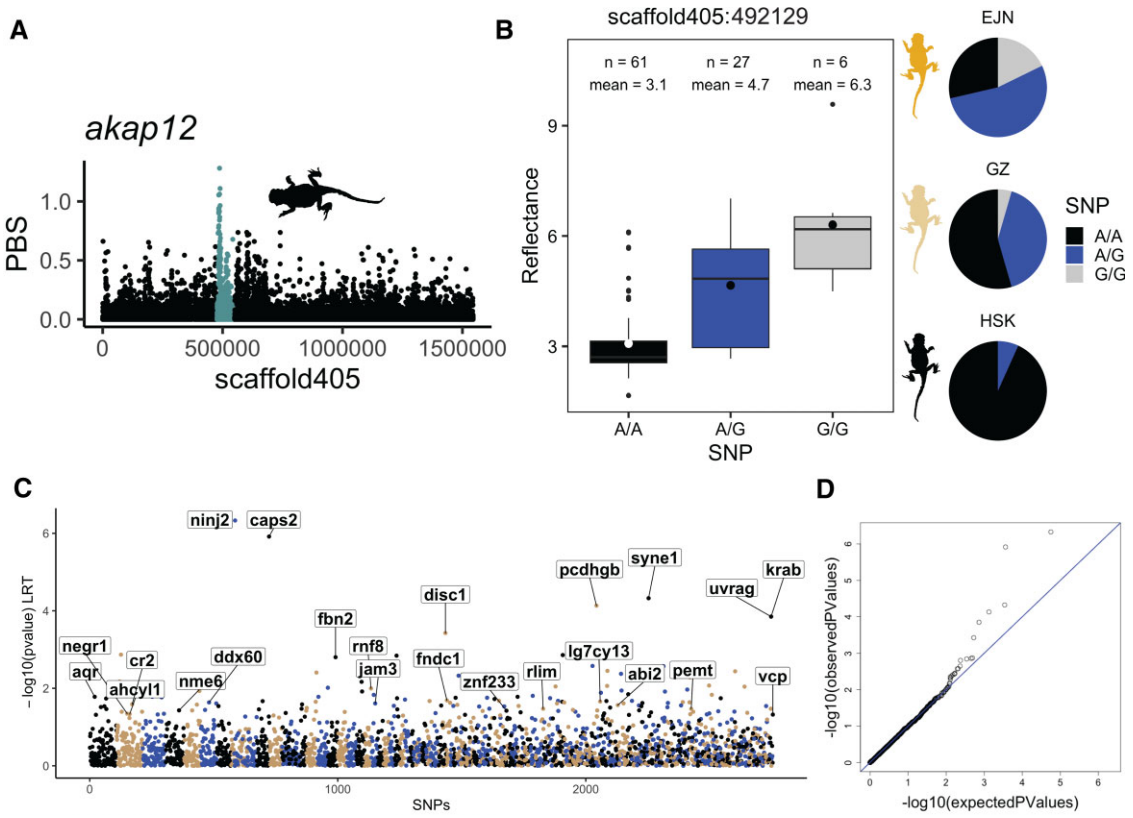

Fig. 4.—Genetic association using reflectance has significant results in *akap12*. (*A*) Single base pair PBS in scaffold405, *akap12* highlighted in teal. (*B*) Reflectance measurements for each genotype in SNP scaffold405:492129 and proportions of genotypes in each population. (*C*) Genomic association using reflectance values. (*D*) QQplot of the genomic association.

of 8.626967 and *P*-value of 0.003312226. We tested a linear model using population of origin and each of these SNPs genotype as explanatory variables and skin reflectance as response variable. The complete results of these tests in all candidate SNPs are shown in supplementary table S9, Supplementary Material online. We would expect the coefficient for the HSK-b melanic population to be significant because it can clearly explain a lot of the reflectance differences (fig. 3*D*). The effect of the genotype at SNP scaffold405:492129, when coded as (0, 1, 2), had a coefficient of 0.7647 with a *P*-value of 6.87e−05, and the effect of belonging to the population HSK-b had a coefficient of −2.0373 and a *P*-value 1.54e−11. Although genotype *P*-value is not significant genome-wide, it is marginally significant, which strengthens the support for the hypothesis proposed by the PBS scan that *akap12* may be related to pigmentation differences between populations. The genotype frequencies in the three populations, and the reflectance value distribution corresponding to each genotype, together contribute to the association of the A allele with lower reflectance values (fig. 3*E*).

We did not find any differences in expression of the *akap12* gene between GZ-y and HSK-b. However, the differential expression analysis we made was only done for the adult stage of the lizard, it is possible that the gene was differentially expressed in earlier stages of development. We also analyzed the sequences from the three different populations and the mutations that we found are all segregating, none of the mutations is fixed. However, it should be noted that with our low-coverage sequencing there is a lot of uncertainty of the genotypes and it is possible that with deeper sequencing fixed mutations in this locus are found. Another possibility is that there are fixed differences in regulatory sequences. Some of the SNPs we found as outliers in the PBS scan are in the akap12 locus but not necessarily in the exons. Color differences have been reported before due to mutations in cis-regulatory elements (Miller et al. 2007).

Analyses of the Yellow Pigment Production Gene slc2a11

After observing signals of selection and differential expression between melanic and nonmelanic morphs, we decided to further investigate the locus containing the *slc2a11* gene. The *slc2a11* gene is involved in leucophore (white

Downloaded from https://academic.oup.com/gbe/article/14/7/evac076/6604964 by University of California, Berkeley/LBL user on 08 July 2022

pigment-producing chromatophores found in fish) and xanthophore differentiation in fish and it is also necessary for yellow pigmentation in birds (Kimura et al. 2014). Five paralogs of the gene were identified in the same region. To examine the possibility of a misassembly artifact, we analyzed coverage with the reads that were used to build the assembly and found it to be even over the different copies (supplementary fig. S4, Supplementary Material online). A dotplot of the region (supplementary fig. S5, Supplementary Material online) revealed that the paralogs of the gene have quite diverged. This divergence can also be appreciated from the gene tree (fig. 3*A*), where paralogs cluster together across populations and phrynocephalus species. Additionally, we also calculated a percentage identity matrix between paralogs and the average identity is 56.70% (supplementary table S4, Supplementary Material online). Thus, this is unlikely to be a misassembled haplotype given that it is well outside of divergence expected for haplotype diversity within species (Leffler et al. 2012). The data indicate it is an ancient duplication. We included individuals from another species from the same genus (*P. vlangalii*) and found the five paralogs present (fig. 3*A*). This is consistent with a very ancient duplication that is also present in close related species. The read coverage for this region was examined in the different populations, but we did not find any signs of structural variation (fig. 3*M*).

There was not a perfect correspondence between the transcripts annotated by Novogene and the paralogs found in the genome, only three of the paralogs were assembled by their de novo assembly of the transcriptome. For this reason, gene expression was reassessed in each of the paralogs (fig. 3*B*). In paralogs 1–4, there is more expression in GZ-y thank in HSK-b. There was almost no expression detected for any of the populations for paralog 2. Paralog 5 has higher expression in HSK-b than in GZ-y. As a validation of the RNA-seq, qPCR was performed and slc2a11 was confirmed to be differentially expressed (supplementary fig. S3*B* and table S8, Supplementary Material online) with higher expression found in GZ-y relative to HSK-b. The qPCR primers were designed for paralog 1, because according to the original analysis of differential expression performed by Novogene, this transcript is significantly differentially expressed.

A gene tree was estimated with a consensus sequence for each paralog using the most common base found in each population and *P. vlangalii* ($n = 10$) (fig. 3*A*). Ortholog sequences of *slc2a11* from four other species were incorporated to the analysis: python (*Python bivittatus*), common box turtle (*Terrapene carolina*), viviparous lizard (*Zootoca vivipara*), and human (*Homo sapiens*). We chose these organisms based on the availability of sequences in NCBI, whereas ensuring a wide range of divergence levels, resulting in a species set including several other lizards, other reptiles, and human sequences. The tree was rooted with the human *slc2a11* gene. One of the paralogs of *slc2a11* forms a separate clade on the tree and is more diverged from the other paralogs than from the same gene in other species (fig. 3*A*). This divergent paralog is the most upstream on the scaffold (fig. 3*L*) and has the lowest sequence similarity with the other four paralogs (supplementary fig. S5, Supplementary Material online). Our sequencing data additionally suggest that paralogs 2–5 are identical between populations.

Since strong signals of selection were detected in this region (fig. 2*E*, *F*, and *H*) and there is differential expression (supplementary figs. S3*A* and S6*B*), regulatory elements in the different populations in this region may have diverged. To address this, we analyzed signals of selection in this region using a sliding 2 kb window, and confirmed that the signals from the genome-wide diversity scans overlap with *slc2a11* paralogs on that scale. Namely, the GZ-y population has increased diversity in this region (fig. 3*G* and *J*) and EJN-y population has reduced diversity (fig. 3*I*). Tajima's *D* is around 0 at this gene in the HSK-b population (fig. 3*E*), whereas it is close to 2 (fig. 3*D*) in GZ-y and -1 in EJN-y (fig. 3*C*). The genome-wide average Tajima's *D* in 2 kb windows in each population is HSK-b: 1.0925, GZ-y: 0.7870, EJN-y: 0.5801. We also plotted the distribution of the Tajima's *D*-values in 2 kb windows and overlay the values of windows in *slc2a11* on top of them (fig. 3*I*), to verify how extreme were these Tajima's *D*-values relative to the genome-wide distribution. The genome-wide average value of GZ-y is larger than 0 so we do not discard the possibility of a bottleneck, however, the values in *slc2a11* are still in the higher tail of the genome-wide distribution. These results could indicate a recent selective sweep in the EJN-y population and possibly balancing selection in the GZ-y population.

We did not find any SNPs in these genes that had significant associations with the reflectance phenotype. We analyzed the sequences from the different populations. We did not find any fixed mutations between populations in paralogs 2–5. We found two nonsynonymous mutations in paralog1 (supplementary fig. S7, Supplementary Material online). One of them is different in the melanic (HSK-b) compared with the nonmelanic populations (EJN-y-GZ-y). We found five mutations between EJN-y and the other two populations, four synonymous and one nonsynonymous.

Discussion

The strongest candidates underlying color variation identified in this study are *slc2a11* and *akap12*. *Akap12* is an important candidate because it was found to be the most divergent gene in the melanic population compared with the nonmelanic populations (figs. 2*C* and 4*A*), having the highest peak in the HSK-b PBS scan. It is a gene that has been reported to have a role in pigmentation phenotypes

Downloaded from https://academic.oup.com/gbe/article/14/7/evac076/6604964 by University of California, Berkeley/LBL user on 08 July 2022

(Santos et al. 2016) and in the α-MSH signaling pathway in *Xenopus laevis* melanophores (Isoldi et al. 2010). We detected a SNP in *akap12* with some evidence of association with reflectance. It is possible that this SNP itself, or by linkage disequilibrium with nearby causal variants, could be responsible for some of the pigmentation differences between melanic and nonmelanic *P. versicolor*.

As a validation of the RNA-seq, the qPCR indicated *slc2a11* was highly expressed in the nonmelanic (yellow) GZ-y population relative to melanic HSK-b (supplementary table S6, Supplementary Material online), which would be consistent with previous studies that reported functions of this gene in lighter/diluted colors. *Slc2a11* was found to be up-regulated in yellow versus black skin fire salamanders (Burgon et al. 2020), establishing it as a good candidate gene for explaining differences between black and yellow morphs of herpetofauna. We found a region in the genome with five divergent paralogs of *slc2a11* that seem to be under selection, and four of them are differentially expressed. Both Tajima's D ($D > 0$) and alpha (increased diversity) in the GZ-y population for this locus are compatible with a balancing selection hypothesis. This population is located near the melanistic HSK-b population, and we hypothesize it could occasionally occupy the dark sand habitat, where the yellow (nonmelanic) phenotype is presumably disfavored or alternatively may receive divergent haplotypes by gene-flow from the melanistic population. Both scenarios involve balancing selection acting through spatially varying selective pressures for the yellow-associated alleles. We would not expect to see a peak in the GZ-y PBS in this region if there is balancing selection in this locus. In contrast, the EJN-y population, located further from dark substrates, shows signatures compatible with a selective sweep ($D < 0$ and decreased diversity). However, we did not discover any direct association with reflectance for this gene.

In TEM analyses on *P. versicolor* skin, we observed different organelle and cell type composition was found between melanic and nonmelanic *P. versicolor* individuals (HSK-b and GZ-y). As expected, melanophores were more abundant in HSK-b overall. Specifically, we observed an apparent increase in HSK-b dermal melanophores, and the unique presence of HSK-b epidermal melanophores. We additionally found clusters of pterinosomes in GZ-y xanthophores, which were not observed in HSK-b. Kuriyama et al. (2013) found most morphs of Japanese four-lined snake to have the same layers of cells: epidermal melanophore, followed by the dermal pigmented layers: xanthophores, iridophores, and melanophores. The exception was the melanic morph which had epidermal and dermal melanophores but lacked dermal xanthophores and iridophores. The authors suggest the melanic morph might be a phenotype that lost the ability to produce xanthophores and iridophores (Kuriyama et al. 2013). A similar hypothesis could be considered for the role of *slc2a11* in melanic *P. versicolor* lizards. Given that *slc2a11*, a gene necessary for xanthophore production (Kimura et al. 2014), was observed as having lower expression in melanic lizards, this may explain the lower proportion of xanthophores or the absence of pterinosomes in melanic *P. versicolor* lizards. We encourage further studies to test the association between xanthophore abundance, *slc2a11* expression, and *slc2a11* genetic polymorphisms in lizards and other organisms.

Materials and Methods

Sampling

Three different populations of *P. versicolor* were sampled (fig. 1*A* and *B*), one with dark dorsal color (melanic) and two with light dorsal color (nonmelanic). The melanic lizard is distributed in a dark substrate (black) in Heishan Kou, China (HSK-b: 95.47°E, 41.00°N, 1,680 m above sea level, $n = 44$), a nearby nonmelanic lizard is distributed in a light substrate (light-yellow/yellow) in Guazhou County, China (GZ-y: 95.61°E, 40.80°N, 1,386 masl, $n = 22$). Approximately 650 km away from the other sampling locations, another nonmelanic population was sampled in the Ejin Banner, Inner Mongolia Autonomous Region (EJN-y: 101.12°E, 42.13°N, 933 masl, $n = 28$). This nonmelanic EJN-y population inhabits a weathered yellow (light) substrate.

Individuals of *P. vlangalii* from Delingha, Qinghai Province (97.23°E, 37.13°N, $n = 10$) were used as an outgroup in some of the analyses. The numbers of individuals described in parentheses correspond to the resequencing sample sizes. Muscular tissue from tail clips was used for DNA extraction for resequencing. A complete list of individuals used for all experiments with their corresponding sex and tissue is provided in supplementary table 10, Supplementary Material online.

Genome Assembly

A female EJN-y individual was sequenced for genome assembly. Illumina BCL files were converted to fastq files using *bcl2fastq*, and genomes were assembled using the Supernova pipeline, version 2.0.0 (Weisenfeld et al. 2017). We subsampled reads using *–maxreads = 800 M*, which yielded a raw coverage of 65.57×. The *pseudohap* output style was used for further analyses. To detect assembly errors, the coverage throughout the genome (supplementary fig. S1*A*, Supplementary Material online) was plotted using a subset of 6.25% of the reads of the female individual that was sequenced for the assembly. We identified large segments with approximately half the coverage and half the heterozygosity of the rest of the genome. Therefore, we concluded that they were likely uncollapsed haplotypes. The coverage of four male and four female lizards was mapped and plotted to check if any half-

Downloaded from https://academic.oup.com/gbe/article/14/7/evac076/6604964 by University of California, Berkeley/LBL user on 08 July 2022

coverage scaffolds seemed to be a sex chromosome. We discarded this possibility because both males and females showed the same coverage across scaffolds (supplementary fig. S2, Supplementary Material online). The scaffolds with half coverage were extracted from the whole genome assembly fasta, to create a fasta file containing only scaffolds with reduced coverage. The assembly fasta file was converted into a BLAST database (Camacho et al. 2009), against which we aligned the half-coverage scaffolds. For alignments with two different scaffolds with more than 99% identity, we deleted the shorter scaffold. To verify that the local alignments made with BLAST fully covered all the scaffolds, we used *MUMmer* (Marçais et al. 2018), which is a tool that allows alignment of whole genomes to other genomes, to find the correspondence between eliminated duplicated scaffolds and remaining scaffolds in the whole genome assembly (supplementary fig. S2, Supplementary Material online). All the code used is available at https://github.com/aguilar-gomez/genome_processing. This processing reduced the original output of supernova from 2.4 to 2 Gb. Scaffolds smaller than 10 kb were excluded for all the downstream analyses, resulting in a genome assembly of 1.7 Gb. We submitted the genome to NCBI to make it publicly available. To match their format requirements, we changed Supernova's gap of 100 Ns to 10 Ns, and removed all leading and trailing Ns of any scaffold. We also removed a few scaffolds (~700 K) that were flagged as possible contamination. The final published genome assembly is 1.6 Gb. We ran *BUSCO 5.2.1* (Simão et al. 2015), which calculates a score of completeness of the assembly based on gene orthology. The lineage parameter was set to "tetrapoda." The score improved after removing duplicated scaffolds and remained the same after we removed the small scaffolds (supplementary table S1, Supplementary Material online).

Genome Annotation

The genome was annotated using *BRAKER2 version 2.1.2* (Hoff et al. 2019). We used the reduced genome assembly (scaffolds <10 kb removed) and 121,482,632 RNA-seq paired reads mapped to this assembly as input with default parameters. RNA was extracted from an EJN-y individual, different from the individual used for the genome assembly, from five tissues (skin, brain, heart, lung, and liver). Total RNA was extracted using TRIzol (Invitrogen, USA) and treated with RNase-free DNase I. RNA quality was determined by Bioanalyzer 2100 using RNA 6000 Nano LabChip Kit (Agilent, USA). Libraries for strand-specific RNA-seq were constructed using TruSeq Stranded mRNA Library Prep (Illumina, USA). Briefly, Oligo(dT) magnetic beads purify and capture the mRNA molecules containing polyA tails. The purified mRNA was fragmented and copied into first-strand cDNA using reverse transcriptase and random primers. In a second strand cDNA synthesis step, dUTP replaced dTTP to achieve strand specificity. The final steps added adenine(A) and thymine(T) bases to fragment ends and ligate adapters. The libraries were subsequently sequenced on the Illumina HiSeq 4000 platform, and 150-bp paired-end reads were obtained. The annotation obtained with *BRAKER2* was used together with an annotation provided by Novogene (supplementary methods, Supplementary Material online) to identify genes in candidate regions detected by the tests described below. We also used nucleotide BLAST searches against the NCBI database to verify each gene ortholog prescription made by *BRAKER2* and the Novogene annotation to confirm gene annotations. The annotation was validated when the same gene was obtained as the first hit.

Measuring the Reflectance Phenotypes

A spectrophotometer (Avaspec-ULS2048CL-RS-EVO-UA), which comprises a mainframe (AvaSpec-2048-USB2), a photo source (AvaLight-DH-S), a reflective probe (FCR-7UV400-2-ME), a probe fixator (RPH-1), and a white diffusing panel (WS-2) used for reflective calibration before measuring, was used to record the skin luminous reflectance (%) of three dorsal measuring points for 78 individuals. These measurements are described in more detail in our previous paper (Tong et al. 2019). Measurements were performed under the same photo source without the influence of natural light. Sampling interval of the spectrophotometer is 0.60 nm, and the reflectance value in the visible range (300–700 nm) of Squamata and their potential avian predators (Bennett and Cuthill 1994) were used for subsequent analysis. The white diffusing panels were used for calibration after preheating the mainframe. Measurements were taken with the probe held at a 90° angle and 2 mm above the lizard's skin. Exposure time was set to 100 ms.

Population Genomics Analysis

We sequenced 44 melanic lizards (HSK-b) and 50 nonmelanic lizards (GZ-y: $n = 22$; EJN-y: $n = 28$) at low-coverage (~3×) (supplementary table S2, Supplementary Material online in Excel format). We kept 30 million di-allelic SNPs where at least 45 individuals had one read with a good mapping score (mapQ 30), read quality score (baseQ 30), and minor allele frequency >5%. This filtering was done with *ANGSD* (Kim et al. 2011; Korneliussen et al. 2014) and *bcftools1.9* (Danecek and McCarthy 2017). Allele frequencies were estimated using *ANGSD* based on genotype likelihoods (Korneliussen et al. 2014). PCA was performed using PCAngsd (Meisner and Albrechtsen 2018), and we calculated the site frequency spectrum (SFS) for each population using *realSFS*.

Downloaded from https://academic.oup.com/gbe/article/14/7/evac076/6604964 by University of California, Berkeley/LBL user on 08 July 2022

Genome-wide F_{ST} and PBS scans between populations (window size = 150 SNPs, nonoverlapping) were performed using *ANGSD*. Previous simulation studies (Crawford and Nielsen 2013) have demonstrated that species with values of $F_{ST} < 0.2$, in general, are highly suitable for adaptive trait mapping using F_{ST} based methods. PBS of the melanic population identifies regions with unusually strong allele frequency differences between melanic versus nonmelanic populations. Outlier peaks were annotated with the genes within the window to which the peak belongs. Outliers were defined as the top 99.5th percentile of windows. Another type of selection scan was implemented using *ANGSD* to estimate θ, the population scaled mutation rate. We used *realSFS* to calculate the folded SFS, followed by *thetaStat* (Korneliussen et al. 2013) to estimate θ using the average number of pairwise differences, in nonoverlapping windows of 30 kb for each population. Then, for each window, we calculated $\alpha_i = \frac{\pi_i}{\pi_i+\pi_j+\pi_k}$ where *I*, *j*, *k* denote the three different populations. A low value of α_I indicates reduced diversity in a region specific to population *I* and may be evidence of selection in species *I*. We also zoomed in on a region of interest using 2 kb windows and used *thetaStat* to calculate Tajima's *D*, π, and α_I. We note that all methods used here are based on genotype likelihoods and do, therefore, explicitly take the genotype uncertainty in 3× sequencing data into account.

Average Genomic Tree

An average genomic tree was calculated using *ngsDist* (Vieira et al. 2015) with genotype likelihoods. We generated 100 bootstrap replicates with a block size of 5 Mb. We then used neighbor-joining (BIONJ), a distance-based phylogeny inference method implemented in *FASTME 2.1.6.2* (Lefort et al. 2015) to infer the trees for the original data and the bootstrap replicates, using subtree pruning and regrafting to improve the tree topology. Finally, *RAxML-NG* (Kozlov et al. 2019) *–support* function was used to collect and summarize the tree, and its bootstrap replicates. To root the tree, we used an individual of *P. vlangallii* as the outgroup.

Differential Expression

Transcriptome assembly, annotation, and differential expression analyses with RNA-Seq (GZ-y = 3, HSK-b = 3) were performed by the company Novogene and are described in the supplementary methods, Supplementary Material online. RNA was extracted from whole dorsal skin from six individuals for the differential expression analysis. The number of reads per library and other details can be found in supplementary table S10, Supplementary Material online. We had the BRAKER annotation but we wanted to complement it with the transcriptome annotation. First, we aligned the transcriptome to the genome using *STARlong v.2.7.0d* (Dobin et al. 2013). The nonprimary alignments were removed using *samtools v.1.10* (Li et al. 2009). The coordinates of the annotated genes were obtained using *bamtobed* from *bedtools v2.27.1* (Quinlan 2014). Transcripts that were annotated as transposons, satellites, and other repeats were excluded. We then focused on the genes that were outliers in any of the three PBS scans, resulting in 1,452 transcripts of interest (supplementary fig. S3*A*, Supplementary Material online). We used the results from DESeq2 (supplementary methods, Supplementary Material online) to identify differentially expressed genes between GZ-y and HSK-b. We selected the log2FoldChange and *P*-value associated with these 1,452 transcripts to calculate *q*-values using FDR only within this subset. Genes that had a corrected *q*-value < 0.05 were marked as significant. Finally, we validated the differential expression of the genes of interest using qPCR.

qPCR Analysis

Dorsal skin tissues were collected from 30 individuals (HSK-b = 15, GZ-y = 15) and the total RNAs were extracted using TRIzol® Plus RNA Purification Kit (Invitrogen id: 12183-555) as described by the manufacturer's protocol. RNase-Free DNase Set (Qiagen id: 79254) was used. cDNA was synthesized using the SuperScript™ III First-Strand Synthesis SuperMix for qRT-PCR (Invitrogen id: 11752-050) following manufacturer's instructions. Real-time PCR was performed with PowerUp™ SYBR™ Green Master Mix (Applied Biosystems, id: 4367659) according to the manufacturer's guides. The cycling program used was 95 °C for 1 min, followed by 40 cycles of amplification (95°C for 15 s, 60°C for 25 s). Relative expressions of target genes were normalized to the housekeeping genes *atcb* and *deef1a1* evaluated by the $2^{-\Delta\Delta Ct}$ method (Livak and Schmittgen 2001) and given as a ratio to control in the experiment. The lizard primers designed based on the assembled genomic sequences and transcript sequences are listed in supplementary table S3, Supplementary Material online.

Genomic Association

Reflectance phenotype data used here were obtained and described in a previous study (Tong et al. 2019). A genetic association with the reflectance phenotype as the response variable was performed in ANGSD with -doAsso 2 (Skotte et al. 2012) using an additive model. This model takes into account genotype uncertainty. We set the parameter -minHigh to 30, indicating that a minimum of 30 highly credible genotypes must be found in a site for the site to be included, this resulted in only 2,754 sites for the test. We used the first ten components of our PCA as covariates. We also performed the association testing using a less

Downloaded from https://academic.oup.com/gbe/article/14/7/evac076/6604964 by University of California, Berkeley/LBL user on 08 July 2022

stringent minH parameter, but the QQplot for those associations had an excess of small *P*-values (supplementary fig. S8, Supplementary Material online). The SNPs where annotated by using bedtools v2.27.1 to intersect the annotation coordinates with the SNP coordinates. SNPs that did not exactly fall within a gene where not annotated. Using this approach, many genes are likely to be missed considering the low SNP density and causal associations might have been missed due to a lack of LD between our marker set and causal variants.

We also used an alternative approach to explore associations in *akap12* and *slc2a11* that used the population of origin as a covariate. We tested different linear models in R to test whether SNPs in the candidate genes were significantly associated with the reflectance phenotypes. We used the following formula: lm (formula = reflectance ~ population + genotype, data = df), where the population of origin (EJN-y, GZ-y, HSK-b) is included as a covariate. The SNPs were selected for this analysis based on high PBS values and *P*-values lower than 0.02 in the genomic association with ten PCs as covariates.

Analysis of the slc2a11 Paralog Region

We generated the consensus sequence from EJN-y individuals using *doFasta* from *ANGSD* (Korneliussen et al. 2014). Then we ran Muscle (Edgar 2004) with the output option set to *clustalw* (Sievers et al. 2011). This generates a summary where we could obtain the percent identity matrix (supplementary table S4, Supplementary Material online). Coverage in this region was calculated on windows of 5 kb, with a slide of 2.5 kb. Reads were mapped using *bwa mem* and coverage was calculated using *samtools bedcov* (supplementary fig. S4, Supplementary Material online). We ran *dotmatcher* using window size of 200 bp and threshold of 30 to generate a dotplot showing the five paralogs of *slc2a11* (supplementary fig. S5, Supplementary Material online). The annotation of the paralogs was confirmed via *blastn* and *blastx* (supplementary table S5, Supplementary Material online).

We also estimated a gene tree of *slc2a11* paralogs. The consensus sequence was generated with *doFasta* from *ANGSD* (Korneliussen et al. 2014), the alignment was performed with *Muscle*, selecting output phylip interleaved. The tree was estimated with *PhyML* (Guindon et al. 2010). Smart model selection was implemented using Akaike Information Criterion. We used BIONJ for tree searching and Nearest Neighbor Interchange for tree improvement. We estimated a pairwise identity matrix between the paralogs by aligning the EJN-y paralogs against each other using *Muscle 3.8.425* (Edgar 2004) with *clustalw* output format. The summary statistic includes this matrix generated by *Clustal2.1* (Sievers et al. 2011).

Expression was reassessed for this region using the manual annotation that we did for the five paralogs (see Genome Annotation). We used *STAR* (Dobin et al. 2013) and *RSEM* (Li and Dewey 2011) to map and quantify the expression level in the region using a gtf file we generated with the exons of this region. The results are shown in fig. 3*B*.

Sequence Analysis of Candidate Genes

We decided to further analyze the sequences of slc2a11 and akap12. For each gene, a consensus sequence was generated for each population using the most common base as described in previous sections. The sequences were analyzed using *Jalview 2.11.1.4* (Waterhouse et al. 2009), using this software we performed alignments with *Muscle* and translated the cDNA into amino acid sequence. We chose the options to wrap and color by percentage identity. We identified all the possible fixed nonsynonymous mutations between populations within the candidate genes.

Transmission Electron Microscopy

Two individuals were used for this experiment, one from each the HSK-b population and GZ-y population. For both individuals we selected a similar central position on the right side of the body flank, collecting samples of the skin outside of the dark stripes found on both populations. The samples were sent to the Electron Microscope Center at Zhejiang University which performed TEM (Hayat 2012; Niu et al. 2019) in a Model JEM-1230 TE Microscope. The detailed protocol is in supplementary methods, Supplementary Material online.

Supplementary Material

Supplementary data are available at *Genome Biology and Evolution* online. The translated abstract (currently available in Chinese and Spanish) can be found at https://github.com/aguilar-gomez/jin_et_al_2022.

Acknowledgments

This work is supported by the National Natural Science Foundation of China (Grant No. 31772447), the Second Tibetan Plateau Scientific Expedition and Research Program (STEP, Grant No. 2019 QZKK0501), and Zhejiang Natural Science Foundation (LY21C040002). We are grateful for the kind help from Yubing Wo in our sampling process. D.A.-G. was supported by the UC-MEXUS CONACYT and the Fulbright García-Robles fellowships, whereas working on this project. This work was initiated by an academic visit of Y.J. to RN's lab supported by the China Scholarship Council. Publication made possible in part by support from

Downloaded from https://academic.oup.com/gbe/article/14/7/evac076/6604964 by University of California, Berkeley/LBL user on 08 July 2022

the Berkeley Research Impact Initiative (BRII) sponsored by the UC Berkeley Library.

Author Contributions

Y.J. and Z.L. collected the specimens and measured the reflectance. Y.J and J.L. did the qPCR. Y.J. prepared the transmission electron microscopy skin (TEM) tissue. T.W. contacted and sent the tissue for TEM. T.A.S. and C.T.M. analyzed the TEM images. Y.J. obtained the sequences. D.A.-G. and D.Y.C.B. assembled the genome. D.A.-G. performed the bioinformatic and statistical analysis, under the advice and supervision of R.N. and P.H.S. D.A.-G. and R.N. wrote the manuscript with comments from all authors. D.Y.C.B. and Y.J. critically revised the manuscript. Y.J. conceived and financed the study. Both Y.J and D.A.-G. contributed equally and have the right to list their name first in their CV.

Ethics Statement

All animal experimental protocols were performed in accordance with guidelines from the China Council on Animal Care and approved by the Ethics Committee of Animal Experiments at China Jiliang University.

Data Availability

We created the BioProject ID PRJNA738137 where we submitted the resequencing data (SRA) (https://www.ncbi.nlm.nih.gov/sra/PRJNA738137). This Whole Genome Shotgun project has been deposited at DDBJ/ENA/GenBank under the accession JAHTLZ000000000 and is part of the same BioProject. The version described in this paper is version JAHTLZ010000000. The RNA-seq data discussed in this publication have been deposited in NCBI's Gene Expression Omnibus and are accessible through GEO Series accession number GSE179069 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE179069).

Literature Cited

Albertson RC, et al. 2014. Genetic basis of continuous variation in the levels and modular inheritance of pigmentation in cichlid fishes. Mol Ecol. 23(21):5135–5150.

Alföldi J, et al. 2011. The genome of the green anole lizard and a comparative analysis with birds and mammals. Nature 477(7366): 587–591.

Bagnara JT, Taylor JD, Hadley ME. 1968. The dermal chromatophore unit. J Cell Biol. 38(1):67–79.

Bennett ATD, Cuthill IC. 1994. Ultraviolet vision in birds: What is its function? Vision Research. 34:1471–1478. doi:10.1016/0042-6989(94)90149-x.

Brejcha J, et al. 2019. Body coloration and mechanisms of colour production in archelosauria: the case of deirocheline turtles. R Soc Open Sci. 6(7):190319.

Burgon JD, et al. 2020. Functional colour genes and signals of selection in colour-polymorphic salamanders. Mol Ecol. 29(7):1284–1299.

Camacho C, et al. 2009. BLAST+: architecture and applications. BMC Bioinform. 10(December):421.

Corso J, Gonçalves GL, de Freitas TRO. 2012. Sequence variation in the melanocortin-1 receptor (MC1R) pigmentation gene and its role in the cryptic coloration of two South American sand lizards. Genet Mol Biol. 35(1):81–87.

Cox CL, Davis Rabosky AR, Chippindale PT. 2013. Sequence variation in the Mc1r gene for a group of polymorphic snakes. Gene 513(2): 282–286.

Crawford JE, Nielsen R. 2013. Detecting adaptive trait loci in nonmodel systems: divergence or admixture mapping? Mol Ecol. 22(24): 6131–6148.

Cuthill IC, et al. 2017. The biology of color. Science 357(6350): eaan0221.

Danecek P, McCarthy SA. 2017. BCFtools/csq: haplotype-aware variant consequences. Bioinformatics 33(13):2037–2039.

Demars J, et al. 2018. New insights into the melanophilin (MLPH) gene affecting coat color dilution in rabbits. Genes 9:430.

Dobin A, et al. 2013. STAR: ultrafast universal RNA-Seq aligner. Bioinformatics 29(1):15–21.

Dorshorst BJ, Ashwell CM. 2009. Genetic mapping of the sex-linked barring gene in the chicken. Poul Sci. 88(9):1811–1817.

Drögemüller C, Philipp U, Haase B, Günzel-Apel A-R, Leeb T. 2007. A noncoding melanophilin gene (MLPH) SNP at the splice donor of exon 1 represents a candidate causal mutation for coat color dilution in dogs. J Hered. 98:468–473.

Edgar RC. 2004. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. BMC Bioinform. 5-(August):113.

Georges A, et al. 2015. High-coverage sequencing and annotated assembly of the genome of the Australian dragon lizard pogona vitticeps. GigaScience 4(September):45.

Gryzińska M, Andraszek K, Jocek G. 2013. DNA methylation analysis of the gene CDKN2B in gallus gallus (chicken). Folia Biologica. 61(3–4):165–171.

Guindon S, et al. 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Syst Biol. 59(3):307–321.

Hasegawa M, Kuriyama T, Brandley M, Murakami A. 2020. Blue, black, and stripes: evolution and development of color production and pattern formation in lizards and snakes. Front Ecol Evol. 8:232.

Hayat M. 2012. Basic techniques for transmission electron microscopy. Elsevier.

Hoff KJ, Lomsadze A, Borodovsky M, Stanke M. 2019. Whole-genome annotation with BRAKER. Methods Mol Biol. 1962:65–95.

Ishida Y, et al. 2006. A homozygous single-base deletion in MLPH causes the dilute coat color phenotype in the domestic cat. Genomics 88(6):698–705.

Isoldi MC, Provencio I, de Lauro Castrucci AM. 2010. Light modulates the melanophore response to α-MSH in *Xenopus laevis*: an analysis of the signal transduction crosstalk mechanisms involved. Gen Comp Endocrinol. 165(1):104–110.

Jin Y, et al. 2020. Dorsal pigmentation and its association with functional variation in MC1R in a lizard from different elevations on the Qinghai–Tibetan plateau. Genome Biol Evol. 12(12): 2303–2313.

Kanetsky PA, et al. 2001. Interaction of glutathione S-transferase M1 and T1 genotypes and malignant melanoma. Cancer Epidemiol Biomarkers Prev. 10(5): 509–513.

Khoo G, Lim TM, Phang VPE. 2012. Ultrastructure of erythrophores and xanthophores of the siamese fighting fish, betta splendens. Isr J Aquac. 64:20620.

Downloaded from https://academic.oup.com/gbe/article/14/7/evac076/6604964 by University of California, Berkeley/LBL user on 08 July 2022

Kim SY, et al. 2011. Estimation of allele frequency and association mapping using next-generation sequencing data. BMC Bioinform. 12(June):231.
Kimura T, et al. 2014. Leucophores are similar to xanthophores in their specification and differentiation processes in medaka. Proc Natl Acad Sci U S A. 111(20):7343–7348.
Korneliussen TS, Albrechtsen A, Nielsen R. 2014. ANGSD: analysis of next generation sequencing data. BMC Bioinform. 15-(November):356.
Korneliussen TS, Moltke I, Albrechtsen A, Nielsen R. 2013. Calculation of Tajima's D and other neutrality test statistics from low depth next-generation sequencing data. BMC Bioinform. 14:289.
Kozlov AM, Darriba D, Flouri T, Morel B, Stamatakis A. 2019. RAxML-NG: a fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. Bioinformatics. 35: 4453–4455.
Kuriyama T, Misawa H, Miyaji K, Sugimoto M, Hasegawa M. 2013. Pigment cell mechanisms underlying dorsal color-pattern polymorphism in the Japanese four-lined snake: pigment cell in color polymorphic snakes. J Morphol. 274(12):1353–1364.
Leffler EM, et al. 2012. Revisiting an old riddle: what determines genetic diversity levels within species? PLoS Biol. 10(9):e1001388.
Lefort V, Desper R, Gascuel O. 2015. FastME 2.0: a comprehensive, accurate, and fast distance-based phylogeny inference program. Mol Biol Evol. 32(10):2798–2800.
Li B, Dewey CN. 2011. RSEM: accurate transcript quantification from RNA-seq data with or without a reference genome. BMC Bioinform. 12(August):323.
Li H, et al. The sequence alignment/map format and SAMtools. Bioinformatics 25(16):2078–2079.
Livak KJ, Schmittgen TD. 2001. Analysis of relative gene expression data using real-time quantitative PCR and the $2^{\Delta\Delta Ct}$ method. Methods. 25(4):402–408. https://www.sciencedirect.com/science/article/pii/S1046202301912629.
Marçais G, et al. 2018. MUMmer4: a fast and versatile genome alignment system. PLoS Comput Biol. 14(1):e1005944.
McLean CA, Lutz A, Rankin KJ, Stuart-Fox D, Moussalli A. 2017. Revealing the biochemical and genetic basis of color variation in a polymorphic lizard. Mol Biol Evol. 34(8):1924–1935.
Meisner J, Albrechtsen A. 2018. Inferring population structure and admixture proportions in low-depth NGS data. Genetics 210(2): 719–731.
Miller CT, et al. 2007. cis-Regulatory changes in Kit ligand expression and parallel evolution of pigmentation in sticklebacks and humans. Cell. 131:1179–1189.
Morrison RL, Rand MS, Frost-Mason SK. 1995. Cellular basis of color differences in three morphs of the lizard sceloporus undulatus erythrocheilus. Copeia 1995(2):397–408.
Nachman MW, Hoekstra HE, D'Agostino SL. 2003. The genetic basis of adaptive melanism in pocket mice. Proc Natl Acad Sci U S A. 100(9):5268–5273.
Niu X-Q, et al. 2019. Mechanism investigation of ethosomes transdermal permeation. Int J Pharm X . 1(December):100027.
Polimanti R, et al. 2015. Haplotype differences for copy number variants in the 22q11.23 region among human populations: a pigmentation-based model for selective pressure. Eur J Hum Genet. 23(1):116–123.
Quinlan Aaron R. 2014. BEDTools: the Swiss-army tool for genome feature analysis. Curr Protoc Bioinform. 47(1):11–12.
Rosenblum EB, Hoekstra HE, Nachman MW. 2004. Adaptive reptile color variation and the evolution of the Mc1r gene. Evolution 58(8):1794–1808.
Rosenblum EB, Römpler H, Schöneberg T, Hoekstra HE. 2010. Molecular and functional basis of phenotypic convergence in white lizards at white sands. Proc Natl Acad Sci U S A. 107(5): 2113–2117.
Santos ME, et al. 2016. Comparative transcriptomics of anal fin pigmentation patterns in cichlid fishes. BMC Genomics 17(September):712.
Senczuk G, et al. 2020. Fifteen shades of grey: combined analysis of genome-wide SNP data in steppe and mediterranean grey cattle sheds new light on the molecular basis of coat color. Genes 11:8.
Sievers F, et al. 2011. Fast, scalable generation of high-quality protein multiple sequence alignments using clustal omega. Mol Syst Biol. 7(1):539.
Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. 2015. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics 31(19): 3210–3212.
Skotte L, Korneliussen TS, Albrechtsen A. 2012. Association testing for next-generation sequencing data using score statistics: association for next-generation sequencing data. Genet Epidemiol. 36(5): 430–437.
Theron E, Hawkins K, Bermingham E, Ricklefs RE, Mundy NI. 2001. The molecular basis of an avian plumage polymorphism in the wild: a melanocortin-1-receptor point mutation is perfectly associated with the melanic plumage morph of the bananaquit, coereba flaveola. Curr Biol. 11(8):550–557.
Tollis M, et al. 2018. Comparative genomics reveals accelerated evolution in conserved pathways during the diversification of anole lizards. Genome Biol Evol. 10(2):489–506.
Tong H, et al. 2019. Effects of substrate color on intraspecific body color variation in the toad-headed lizard, phrynocephalus versicolor. Ecol Evol. 9(18):10253–10262.
Vieira FG, Lassalle F, Korneliussen TS, Fumagalli M. 2015. Improving the estimation of genetic distances from next-generation sequencing data. Biol J Linn Soc Lond. 117(1):139–149.
Wasmeier C, Hume AN, Bolasco G, Seabra MC. 2008. Melanosomes at a glance. J Cell Sci. 121(Pt 24):3995–3999.
Waterhouse AM, Procter JB, Martin DMA, Clamp M, Barton GJ. 2009. Jalview version 2—a multiple sequence alignment editor and analysis workbench. Bioinformatics 25(9):1189–1191.
Weisenfeld NI, Kumar V, Shah P, Church DM, Jaffe DB. 2017. Direct determination of diploid genome sequences. Genome Res. 27(5):757–767.

Associate editor: Bonnie Fraser

Downloaded from https://academic.oup.com/gbe/article/14/7/evac076/6604964 by University of California, Berkeley/LBL user on 08 July 2022