A deep learning phenotyping method for genetic analysis of 3D micro-CT data

A. Karshenas[1,3*], T. Linderoth[1,4], R. Zatha[2], B. Rusuwa[2], and R. Durbin[1*]

[1]Department of Genetics, University of Cambridge, Cambridge, UK
[2]Chancellor College, University of Malawi, Zomba, Malawi
[3]Current address: Biophysics Graduate Group, University of California, Berkeley
[4]W.K. Kellogg Biological Station, Michigan State University, Hickory Corners, USA

* Corresponding authors. Email: ak2272@cantab.ac.uk (AK) rd109@cam.ac.uk (RD)

Abstract

The number of Genome-Wide Association Studies (GWAS) has been growing rapidly in recent years due to developments in genotyping and sequencing platforms. When applied to quantitative traits, these and other statistical genetics approaches require large amounts of consistently and accurately measured phenotypes. Here, we introduce a computational toolbox based on deep convolutional neural networks that we have developed to phenotype quantitative traits describing morphology from micro-CT-scan image datasets. We illustrate the use of this Deep Learning Phenotyper (DLP) on a sample set of craniofacial CT scans of 118 samples from two very closely related species of Lake Malawi cichlid fish, *Maylandia zebroides* and *Cynotilapia zebroides*. We show that the pipeline constructed and implemented here is capable of measuring morphological skeletal phenotypes with high accuracy. We also demonstrate how this pipeline can be integrated with existing GWAS frameworks to identify candidate association loci. We believe the methods we present here will be valuable for groups studying quantitative morphological traits not only in fishes, but in other vertebrates using CT scan datasets.

Keywords: phenotyping, deep learning, morphological analysis, quantitative genetics, GWAS, micro-CT data

Running headline: Deep learning phenotyper for micro-CT data

1 Introduction

One of the most striking but complex ways that individuals vary within and between species is in their three-dimensional shape, or morphology. The genetic basis of morphological variation has been studied in many vertebrates beyond humans in the last few decades (Roff and Mousseau, 1987; Wallace et al., 2014). Although landmarking and other manual or semi-manual methods exist for quantification of morphology using live samples, photographs, and three-dimensional computer tomographic (CT) images , they are hard to scale to the large datasets that are required for statistical genetic analyses (Webster and Sheets, 2010). In this work, we introduce a deep learning-based, three-dimensional phenotyping method that can be used to scale up and automate morphometric measurements from CT scan data.

Since the development of tomography methods and the first CT scan of a living human brain in 1971 (Hounsfield, 1973), CT scanning instruments have become globally available, and are widely used in medical diagnosis, rapid use of CT scans in COVID-19 diagnosis being a perfect example (Li and Xia, 2020), as well as for other purposes. In particular, they have been used to generate high-resolution images of skeletons and soft tissues in many vertebrates for morphometric research (Steger et al., 2012; Khmelinskii et al., 2012). Raw CT images are generated through a rotating pair of conventional X-ray generators and detectors, that provide a stack of x-ray projections captured from different orientations. These images generated by the scanners then go through tomography algorithms in order to generate CT images (Withers et al., 2021). These methods back calculate the density of materials based on x-ray absorption (Hounsfield, 1980), resulting in high-resolution three-dimensional images that are commonly stored as stacks of two-dimensional (2D) images, typically ordered along the anterior-posterior axis of the body, giving radiologists and researchers access to detailed images of internal organs and skeleton structure.

These CT methods have started to generate large datasets across samples that can be used with genetic data to uncover causal links in phenotypic variation observed both within and across species. Classic examples of such studies are Genome-Wide Association Studies (GWAS) (Frary et al., 2000; Burton et al., 2007; Wood et al., 2014; He et al., 2017; Buniello et al., 2018; Soliai et al., 2021) and heritability analysis (Visscher et al., 2008). While genetic analysis of simple morphological traits (such as human height) is routine (Wood et al., 2014), there is a desire to apply the GWAS technology to more complex multi-dimensional phenotypes. The power of these computational approaches increase with the sample size (Spencer et al., 2009; Hong Eun Pyo, 2012),making accurate, high-throughput phenotyping pipeline desirable. Hence, our aim was to develop an unbiased, reliable, and easy-to-use toolbox that can be applied, with minor adjustments, to CT scans of most vertebrates that is capable of generating large phenotypic datasets.

We demonstrate here the application of this pipeline to CT-scans of cichlid fishes from the closely related species *Maylandia zebroides* and *Cynotilapia zebroides*, which exhibit variable morphology in the jaw and head. In section 2 we first introduce this dataset and then describe the methods used in the software, with results summarised in section 3. The traits measured using the method were used to identify candidate genetic loci underlying craniofacial variations which we discuss in sections 3 and 4.

2 Materials and Methods

2.1 Malawi cichlid dataset

Lake Malawi, the southernmost great lake in the East African Rift (Ebinger, 2005), is home to one of the most diverse vertebrate radiations in the world, with more than 500 endemic cichlid fish taxa that all are believed to have diverged within the last million years (Johnson et al., 1996; Malinsky et al., 2018; Svardal et al., 2019). Despite this short diversification time, the morphological and ecological variation between species is extensive (Kornfield and Smith, 2000), and the mechanisms through which this diversity has emerged have been the subject of many studies (Coyne, 1992; Schluter, 2009; Malinsky et al., 2018; McGee et al., 2020; Nosil et al., 2021). However, while there has been increasing use of large scale sequencing data in these studies (Malinsky et al., 2018; McGee et al., 2020) so far few analyses have combined dense genetic data with detailed quantitative morphology information (Ronco et al., 2021).

Here we focus on two closely related species of rock-dwelling cichlids from the Mbuna group of species in Lake Malawi: *Maylandia zebroides* and *Cynotilapia zebroides*. These are superficially similar (Figure 1 (A)), but have different adult teeth and exhibit more subtle differences in other features, in particular in their jaw and craniofacial structures. An attractive approach to obtain highly precise quantitative measurements of these features is micro-CT scanning, which can give exquisitely detailed 3D information on the underlying skeletal structure (Figure 1 (B)).

In total a set of 118 (115 after removal of 3 outliers) fish from four locations were used for this study that were collected in two field trips to Lake Malawi in 2016 and 2017 (see Table Supplementary 1 for details). These were fixed in formaldehyde and ethanol, and scanned at the Cambridge Biotomography Centre using a Nikon XT 225 ST micro-CT scanner. Tomographic reconstruction from the raw images generated stacks of 2D images with a separation of 10μm. Whole genome shotgun DNA sequencing data were also collected on an Illumina HiSeq X instrument.

2.2 Pre-processing

The Malawi cichlid CT image dataset consists of stacks of 2D images that are re-constructed with a uniform separation specified in the metadata files generated by the scanner. These tomography maps are commonly presented in the anterior-posterior (AP) projection, however, they can be used to generate three sets of image stacks for each sample, one for each projection: Left-Right (LR), Anterior-Posterior (AP), and Dorsal-Ventral (DV) - as shown in Figure 1 (B). These images are commonly stored as 16-bit tiff files and can easily occupy tens of gigabytes per sample (mean $24GB$ per sample for our dataset). The orientation of the body, amount of blur, and noise are also all dependent on scanner settings and can vary significantly between scans taken on different days. An example of such imperfections in the scans is shown in Figure 2 (C) with the noise highlighted in red. In order to address these issues, we first implemented a pre-processing pipeline, highlighted in red in Figure 2 (A), to compress, denoise, and normalise all images in the dataset.

2.2.1 Compression

Image stacks are first compressed using a dynamic bit-wide compression algorithm that truncates bits showing random behaviour when looking at the ensemble of pixels in each of the images (Figure 2 A). The degree of randomness of each bit is determined using the frequency of 1s as well as the spatial correlation of the bit with the same bit of neighbouring pixels (see Section 4 for a description on how these spatial correlations were computed). Figure 2 (B) shows how the frequency of ones and

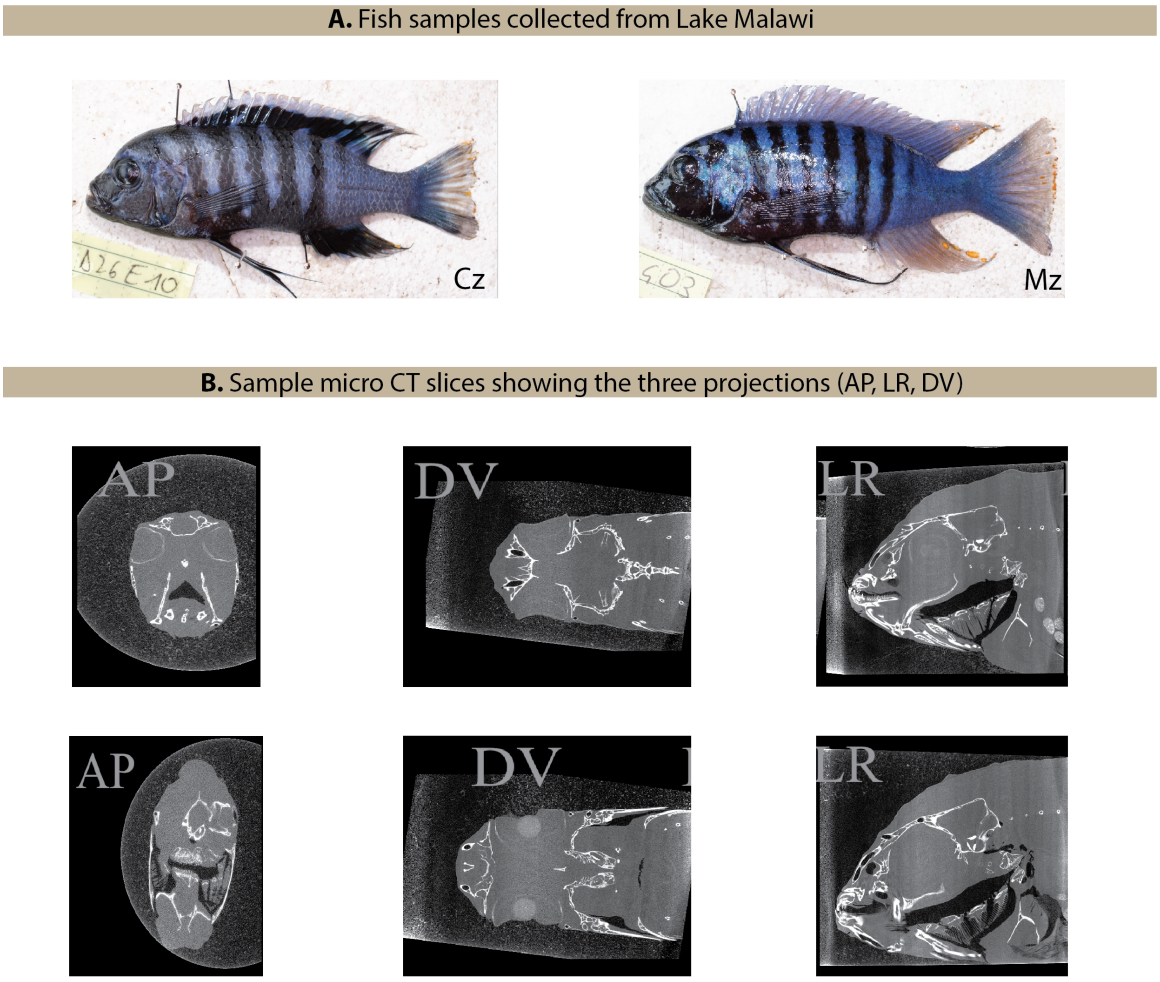

Figure 1: (A) Images of individual *Maylandia zebroides* (MZ), and *Cynotilapia zebroides* (CZ) fish. (B) Sample CT scans of *Maylandia zebroides* individuals perpendicular to the three primary axes: Anterior-Posterior (AP), Dorsal-Ventral (DV), and Left-Right (LR)

spatial correlations of bits vary as a function of bit depth. The compression method implemented in the pre-processing pipeline compresses images by truncating bits (shown in red in Figure 2 B) with a compression ratio of 10-12 depending on bit profiles. Images are then stored using lossless Joint Photographic Expert Group (JPEG) format for further lossless compression (Matsuda et al., 2005).

2.2.2 Noise reduction

Compressed images are then passed through a low-pass Fourier filter implemented as a fifth-order Butterworth filter (Butterworth et al., 1930) to suppress any white additive noise at frequencies higher than the image bandwidth (50 Hz), determined empirically by investigating ensembles of Fourier transforms. While white additive noise can be removed using linear Fourier filters, aberrations that are spatially correlated can not be removed, such as those highlighted in red in Figure 2

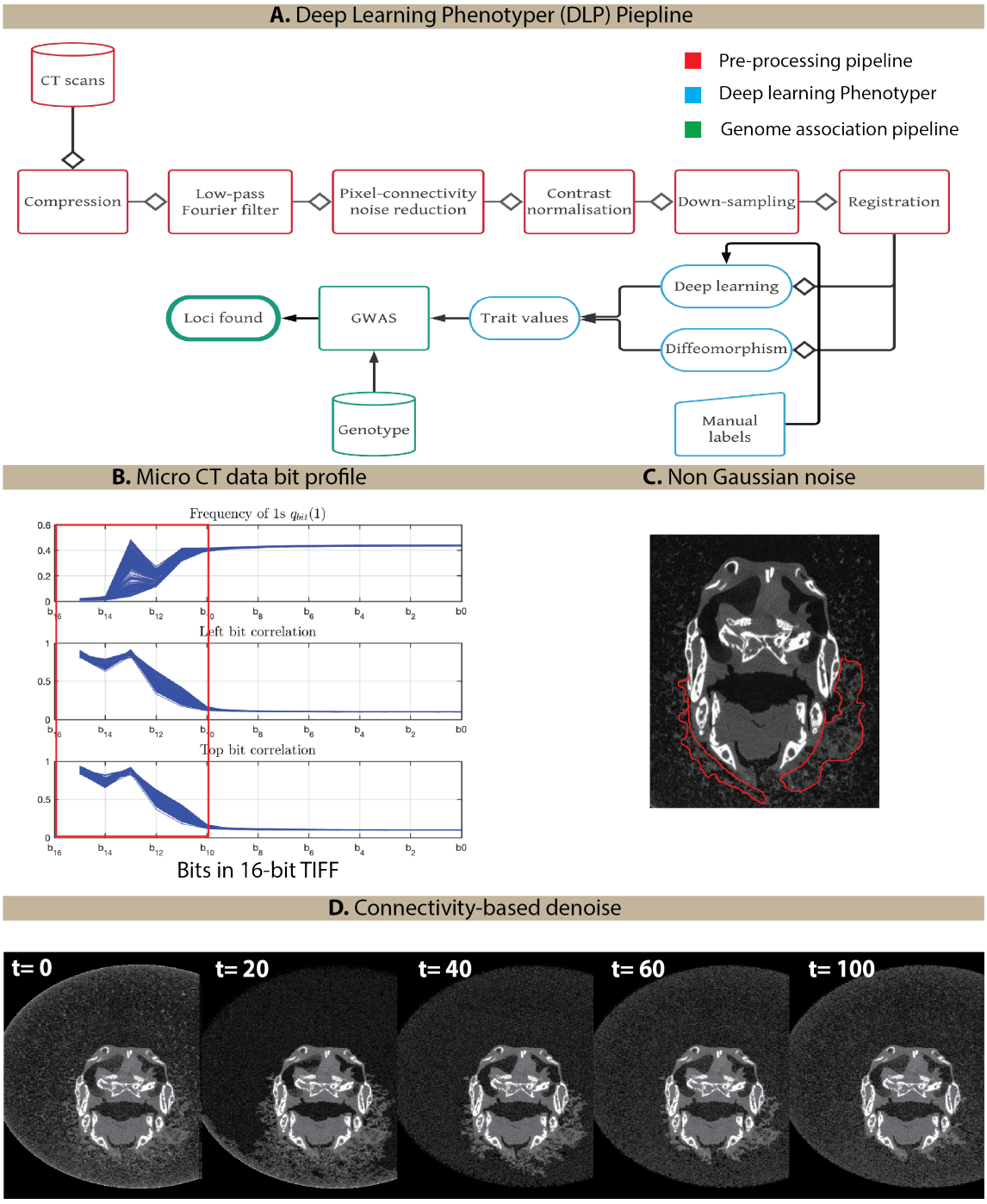

Figure 2: (A) Deep learning phenotyping pipeline with pre-processing steps shown in red. (B) Variation in the frequency of "1s" in bits as well as spatial correlation plotted as a function of bit depth with the red window showing the least significant bits that can be truncated with minimal information loss. (C) Sample CT scan showing non-additive, spatially correlated aberrations in red. (D) Image stack showing how the efficacy of the pixel connectivity-based filter varies with the choice of binary thresholding value $t \in \{0, 20, 40, 60, 100\}$) from left to right.

(C). A pixel-connectivity denoising filter was, therefore, implemented to remove these artefacts based on the contrasts between pixels with six-connectivity in three dimensions. Algorithm 1 describes how this filter was implemented.

Algorithm 1 Connectivity-based denoising

1: **for** $t \leq L$ **do**
2: Create a binary mask $M(i,j) = 1(I(i,j) \geq T(t))$
3: Compute a 6-pixel connectivity matrix $CC(i,j)$ that counts the number of connected neighbours that are above the threshold $T(t)$.
4: Mask CC with M to ignore pixels that are not locally similar.
5: Mask the original image I with the updated mask to remove the superficial reflections.
6: **end for**

The value of t was iterated over a range of pre-determined values stored in an array T of length L. An optimal value of 60 was chosen by empirically investigating a series of 100 images. Figure 2 (D) shows how the efficacy of this filter in removing the aberrations varies with the choice of binary mask thresholding value.

2.2.3 Contrast normalisation and down sampling

Denoised image stacks are passed through a contrast normalisation in order to normalise the contrast between images of samples scanned with different scanner settings specified in the *xteckt* files. The normalisation process is done through histogram matching, with one of the samples selected as a reference. The contrast normalisation is then followed by a Gaussian-based convolutional down-sampling step in order to normalise the dimension of all image stacks from different samples. These pre-processing methods are crucial to ensure that the variations registered using the phenotyping platform are not due to differences in scanner setup and are true measurable variations between samples. Images passed through this pre-processing pipeline were then used for phenotyping as shown in blue in Figure 2 (A).

2.3 Deep learning 3D phenotyper

We adopted a transfer-learning classification-based approach to phenotype morphological features from the image stacks pre-processed as discussed above. Our strategy was to train classifiers for whether or not a 2D slice of the data contained a 3D feature, such as the eye or pharyngeal jaw, and then use these to measure a bounding box for the feature with lengths along each of the three axes as in 3 (A). We call this system the Deep Learning Phenotyper (DLP).

To train the classifiers, a manually labelled training dataset of 10,000 2D images was generated for each feature and for each of the three standard dimensions, by selecting fish randomly and assigning blocks of slices that contained the feature to the set until a sufficient number of images were obtained. These training sets were then randomly shuffled and an equal number of negative images were selected at random from the remaining slices in the same fish. Labelled images were then fed into the following pre-trained networks: *Alexnet*, *Resnet50*, *Resnet18*, and *GoogleNet* (Krizhevsky et al., 2012; Szegedy et al., 2014; He et al., 2015b,a) respectively. Multiple feature spaces were generated using different pooling layers of these networks, and were then used to train a linear Support Vector Machine (SVM) (Boser et al., 1992) using the manually annotated images. Evaluation on a held out sample of labelled images identified the last average pooling layer of *Resnet50* as the optimal architecture for the task, which was consistent with previous studies showing high efficacy

of *Resnet50* in CT datasets (Peng et al., 2020). The pre-trained *Resnet50* network was therefore used to generate a 2048-dimensional feature vector for every image in the image stack for individual fish scans along a given projection. Figure 3 (B) shows the first two Principal Components (PCs) of the training feature space for a binary classifier of an eye/ no-eye classification.

We next evaluated a range of SVM kernel functions for classification over this feature space, using 60% of the annotated dataset for training (7,500 images) and 20% for model selection (2,500 images). Using this approach, a linear SVM model was shown to outperform all the other models for all three projections (AP, LR, and DV). The confusion matrices and classification accuracy given in the Results section were then calculated using the final 20% of the evaluation data which had not been previously used in the training and optimisation process. Images that are not labelled are first passed through *Resnet50* to encode them each into a 2048-dimensional vector. These vectors are then passed through the trained SVM models to classify each image, with the output providing a posterior probability of the slice containing the major craniofacial feature that the SVM was trained on. These values are then plotted along each of the three projections in order to create "projection profiles" that indicate how long, wide, and deep each of the labelled anatomical features is (Figure 3 C).

Due to noise and other adverse effects, the SVM classifier can have uncertain posterior predictions and even misclassify some of the images (Figure 3 C). To address this, the posterior probabilities in a profile are passed through a smoothing algorithm that leverages the continuity of anatomical features to correct misclassified images, as shown in Algorithm 2. This algorithm also returns the left- and right-most positions at which the posterior probability is above the 60% threshold. These boundaries are used to calculate a potential error measure for the feature. Figure 3 (D) shows the output of this algorithm applied to the initial SVM output shown in Figure 3 (C).

Algorithm 2 Classification error removal

1: Specify a maximum window size W.
2: Initialise $P(i)$ to be 1 if the SVM posterior of image i,$p(i)$ is greater that 0.6 else 0 ($\forall i \in \{1, 2, \ldots, L\}$)
3: **for** $i \leq L$ **do**
4: Set $w = \min(W, i, L - i)$ to ensure the window does not extend beyond the edge
5: Set corrected predicted $CP(i) = \text{mode}(P(i - w : i + w)))$
6: **end for**
7: Return the corrected prediction vector CP.

The final step in the pipeline (Figure 2 A) is a conversion step that maps the number of consecutive images labelled to contain the same anatomical feature to the physical length of that feature along the current axis by multiplying by the voxel size along that axis.

2.4 Genetic analysis

We identified candidate genetic loci associated with variation in the measured traits of 61 *M. zebroides* and 54 *C. zebroides* (after removing three outliers as described below) using the generalized linear modeling framework of ANGSD (Korneliussen et al., 2014). This approach accounts for genotyping uncertainty through incorporation of genotype posterior probabilities in the association likelihood model. Following standard quantitative genetics approaches, ANGSD models the traits $\mathbf{y} = (\mathbf{y}_1, ..., \mathbf{y}_{115})$ of the 115 individuals as normally distributed, such that for the i'th individual $\mathbf{y}_i \sim N(\alpha \mathbf{x}_i^\top + \beta \mathbf{g}_i, \sigma^2)$. The genotype $\mathbf{g}_i \in \{0, 1, 2\}$ denotes the number of minor alleles at a given SNP and $\mathbf{x}_i = (1, \mathbf{x}_{i1}, ..., \mathbf{x}_{i8})$ is a vector of covariates that control for variables that may confound

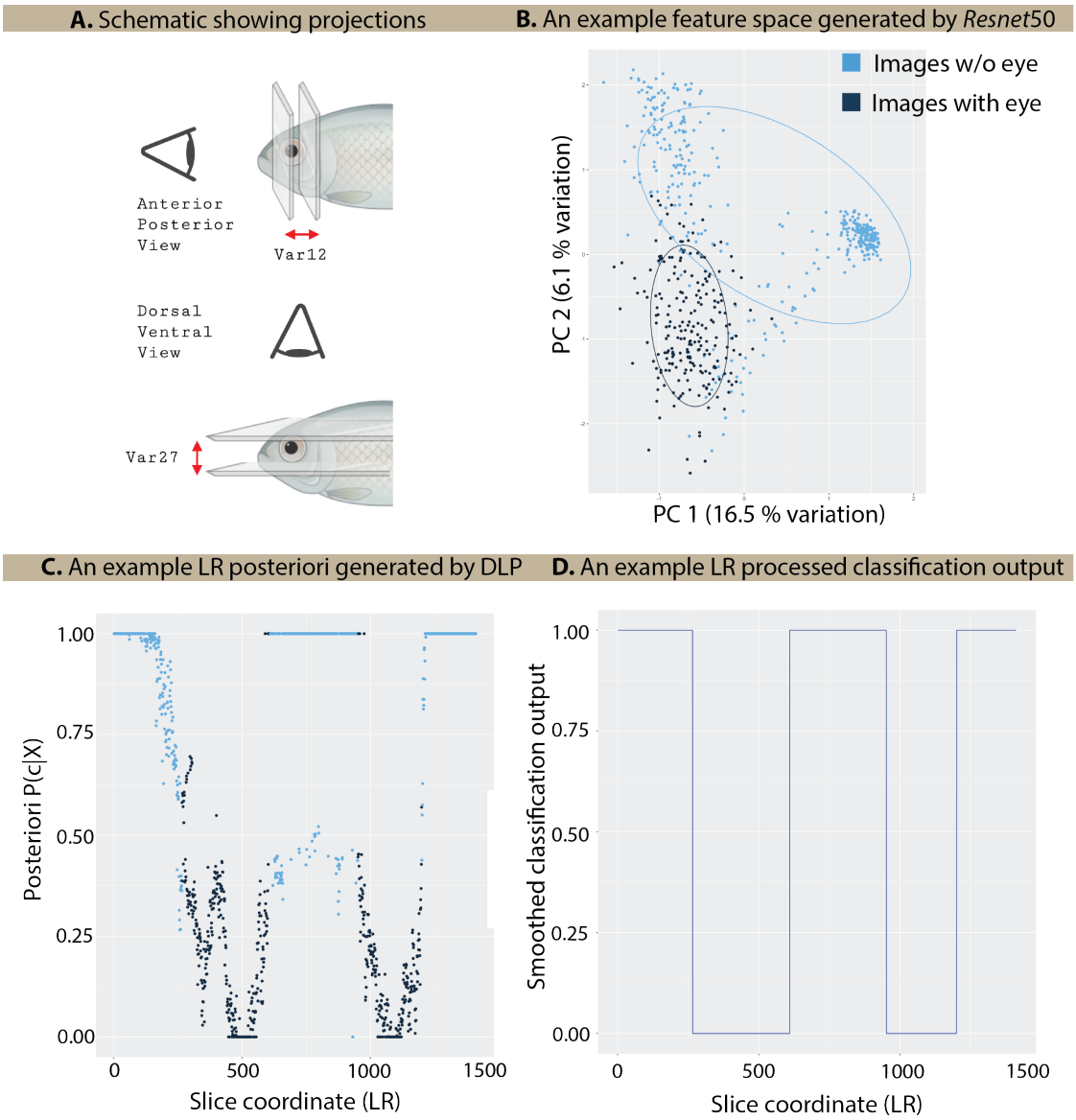

Figure 3: (A) Schematic diagram showing how classification of major anatomical feature (e.g the eye) can be used to determine how wide, deep and high that anatomical feature is using all three projections (AP,LR, and DV). (B) The first two Principal Components (PCs) of the LR feature space generated by the last average pooling layer of *Resnet50*. The colours correspond to different labels on the images (i.e eye or no eye classes) - (C) Posteriori $P(c|X)$, generated by the DLP for images in the LR projection of a sample fish, plotted against the LR axis coordinate of the body. (D) Label vector filtered for abnormalities in classification using the smoothing algorithm given in Algorithm 2

causal genetic associations. We used the length of the head ($\mathbf{x}_{i1}$) to account for the overall size of the fish and scores ($\mathbf{x}_{i2}, ..., \mathbf{x}_{i8}$) of the first seven principal components (PCs) of a principal component analysis (PCA) of genome-wide genetic variation to control for population structure. We fit global maximum likelihood estimates of $\alpha = (\alpha_0, \alpha_1, ..., \alpha_8)$, and then for each locus compared the maximum likelihood with β free, allowing for uncertainty in the genotypes using the genotype posterior probabilities, to the likelihood when $\beta = 0$, which corresponds to the null hypothesis of no genetic effect, using a likelihood ratio test (Skotte et al., 2012).

We tested for genetic trait associations at 2,693,904 biallelic SNPs having a minor allele frequency (MAF) of at least 2% and for which at least 10% of individuals had a posterior probability of 0.9 or greater for at least two of the three possible genotypic configurations (0, 1, or 2 minor alleles). We also required that at least five individuals had a called genotype with quality of at least 15 using GATK (McKenna et al., 2010), and that at least two of the three genotypic configurations were represented in this manner at each tested SNP. We calculated a p-value of genetic association for each SNP by comparing twice the log likelihood ratio to a χ^2_1-distribution, which we considered significant if it exceeded 5% after a conservative Bonferroni correction for the number of tested SNPs (-log10(p-value) > 7.73).

The PCA of genome-wide genetic variation used to control for population structure in the GWAS was performed using ngsTools (Fumagalli et al., 2014) and R. Genotype likelihoods for all individuals calculated with ANGSD were used with ngsTools to estimate a genetic covariance matrix from SNPs with a likelihood ratio test p-value for being variable below 1e-6 and minimum MAF of 5%. This covariance matrix was decomposed using the 'eigen' function of R, yielding the eigenvectors used to control for population structure. We used the first seven PCs as covariates in the GWAS since these captured the genetic structure between the species and sub-populations within species (see Figure 5 B for the first two PCs).

2.5 Implementation and availability

The main package for DLP and pre-processing is written in MATLAB 2021R, with ancillary scripts in R and bash. All of the code required to implement DLP and also the R scripts used to generate the plots in Figures 3, 4, and 5 are available along with documentation at http://github.com/armankarshenas/DLP. This documentation guides users through pre-processing images before labelling and training the SVM. There is also guidance and scripts for users to integrate the package with ANGSD if they seek to run genetic association tests.

3 Results

3.1 Phenotyping method

Our goal was to implement an automated package that would require minimal manual input to ease the phenotyping process for making skeletal measurements from micro-CT data, while keeping a high level of accuracy. To this end, we first explored three diffeomorphic mapping algorithms (Ashburner and Friston, 2000; Friston et al., 2007; Ashburner, 2007) to register variations in morphology to a reference image (Figure 2 (A)). However, none of these were successful, we believe due to not having an adequate reference atlas as well as noise and aberrations in the scans. The variational methods we investigated had previously been applied to Magnetic Resonance Imaging (MRI) scans (Friston et al., 2007), for which high-resolution atlas datasets exist and all input images are extensively manually pre-processed.

We next considered deep learning methods, due to their ability to generalise well when trained with care to avoid over-fitting. Recent developments in image classification and segmentation using deep learning have resulted in comparable performance to manual classification and segmentation (LeCun et al., 2015; Mukhoti and Gal, 2018; Grigorescu et al., 2019). Within deep learning methods, we first attempted to extract phenotypic measurements of the skull by segmenting the CT images, however, the performance of segmentation implemented using a variety of deep convolutional neural networks (Havaei et al., 2015; He et al., 2015a; Chen et al., 2018) was poor. We then investigated architectures developed for object detection(Fernández et al., 2007; Redmon et al., 2015; Redmon and Farhadi, 2016, 2018) to localise biologically significant bones, fragments or other anatomical features within the skull. While the performance improved over segmentation approaches with the same level of manual training information, the overall measurement accuracy was still not acceptable, with under 50% overlap between the predicted bounding boxes and the manually annotated ones in the evaluation dataset.

To resolve these issues, we decided to formulate the phenotyping problem as a classification task. More specifically, we used a classification network to assess whether or not a given anatomical feature was present in a given two dimensional slice of the data. By applying this to all the slices along an axis, for example the Anterior-Posterior (AP) axis, we could identify intervals along that axis that contain the feature, and hence measure its length along that axis. By carrying out this procedure on image stacks generated for each of the AP, Left-Right (LR) and Dorsal-Ventral (DV) axes we could measure the size of features such as bones in each of the three canonical orthogonal directions. Figure 3 (A) shows an example of how classification can be used in two different projections to measure the eye diameter of a fish.

This classification approach allowed us to use transductive transfer learning approaches (Pan and Yang, 2010; Weiss et al., 2016) that are commonly implemented using very well-established networks (Krizhevsky et al., 2012; Szegedy et al., 2014; He et al., 2015b) , trained on very large annotated image datasets (see (Lin et al., 2014; Russakovsky et al., 2015) for examples of these datasets). Many different network architectures have been proposed in the last few years with astonishingly high accuracies in image classification, some beyond human performance. It is common lore that the first layers of deep CNNs trained on large datasets extract morphological features that are essentially independent of the image dataset used. This suggests that a much less complex model could be trained if only the final few layers of these deep networks were to be trained on the new dataset, as opposed to training millions of parameters from scratch.

We implemented this approach as described in detail in the Methods section 2.3. There are two steps in how we extract phenotypic data from images using CNN classifiers. The first step trains a support vector machine (SVM) on the features generated by the last pooling layer of the *ResNet50* network (He et al., 2015b). Figure 4 panels (A,B,C) show confusion matrices for classification of slices by SVMs for each of the AP, LR and DV axes, evaluated on held out labelled data using a 50% posterior probability threshold. These give overall SVM classification accuracies of 94%, 92% and 89% for the three axes respectively. A second step smooths these posterior estimates across slices to give a continuous interval along the axis that is predicted to contain a given feature (as in Figure 3 D), then converts the length of this interval into a physical length using the voxel spacing. Training was fast, taking approximately ten minutes on a single cluster node with four GPU cores.

3.2 Craniofacial trait measurement

The pipeline described above was applied to the dataset of 61 *M. zebroides* and 54 *C. zebroides* micro-CT scans (see Methods section 2.1). Twenty-one different measurements were made as listed

in Table 1. Nineteen of these are quantitative craniofacial traits, while two measures, Var17 and Var23, correspond to the maxima of the error measures across all features for the LR and DV projections.

Figure 4 (D) shows the size range of measurements for each feature. The error in measurement was assessed by comparing the DLP measurements with measurements that were conducted manually using the Slicer interactive software (Kikinis et al., 2014). Figure 4 panels (E) and (F) show these error values both in absolute mm and relative to the size of the respective measurement. Absolute errors are typically below 0.1mm, mostly within 1% of the trait value. This is substantially below the standard deviation of trait values seen across individuals, as seen in Figure 4 (D) and Table 1, indicating that the automated DLP method generates informative measurements for downstream analysis of inter-individual variation. We note that the error values presented here, which are several times the slice resolution of 10μm are upper estimates, because there is likely to be error in the manual measurements using Slicer.

Table 1: Description of the traits extracted by DLP with their summary statistics.

Trait	Description / Projection	Mean (mm)	Standard deviation (mm)
Var8	Length of the frontal jaw / AP	6	1.34
Var9	Length of the Preorbital bone / AP	13.0	2.13
Var10	Length of the Posttympanic bones / AP	16.4	2.46
Var11	Total length of the head / AP	19.4	2.91
Var12	Eye diameter / AP	6.69	1.39
Var13	Length of the Posttympanic bone / AP	10.1	1.45
Var14	Eye to back of head separation / AP	13.1	1.92
Var15	Length of the pharyngeal jaw / AP	3.44	0.87
Var16	Pharyngeal jaw to back of head separation / AP	6.45	1.23
Var17	Error (resolution) in the LR projection	4.18E-5	1.45E-5
Var18	Depth of the first eye / LR	3.11	0.57
Var19	Width of the Preorbital bone - left / LR	7.98	1.46
Var20	Total width of the head / LR	11.1	1.99
Var21	Length of the pharyngeal jaw / LR	4.87	1.07
Var22	Width of the Preorbital bone - right / LR	8.00	1.58
Var23	Error (resolution) in the DV projection	4.36E-5	1.45E-5
Var24	Height of the Frontal bone / DV	6.60	1.70
Var25	Height of the Preorbital bone / DV	11.0	2.15
Var26	Total height of the head / DV	23.3	3.96
Var27	Depth of the eye / DV	4.44	0.99
Var28	Height of the Preopercle bone / DV	16.7	3.10

3.3 Illustrative use for quantitative genetics

To illustrate the utility of the DLP measurements for downstream analysis, we present here an initial GWAS analysis of these samples using general linear models. Figure 5 (A) shows the first two principal components (PCs) of the nineteen traits, colour coded by species. It is apparent that PC1 largely separates the samples by species. PC2 separates out three samples from the rest, which we treat as outliers and removed from the association tests. We note that in general the *M. zebroides* samples are larger than the *C. zebroides* samples, so that all measurements tend to be larger, although the

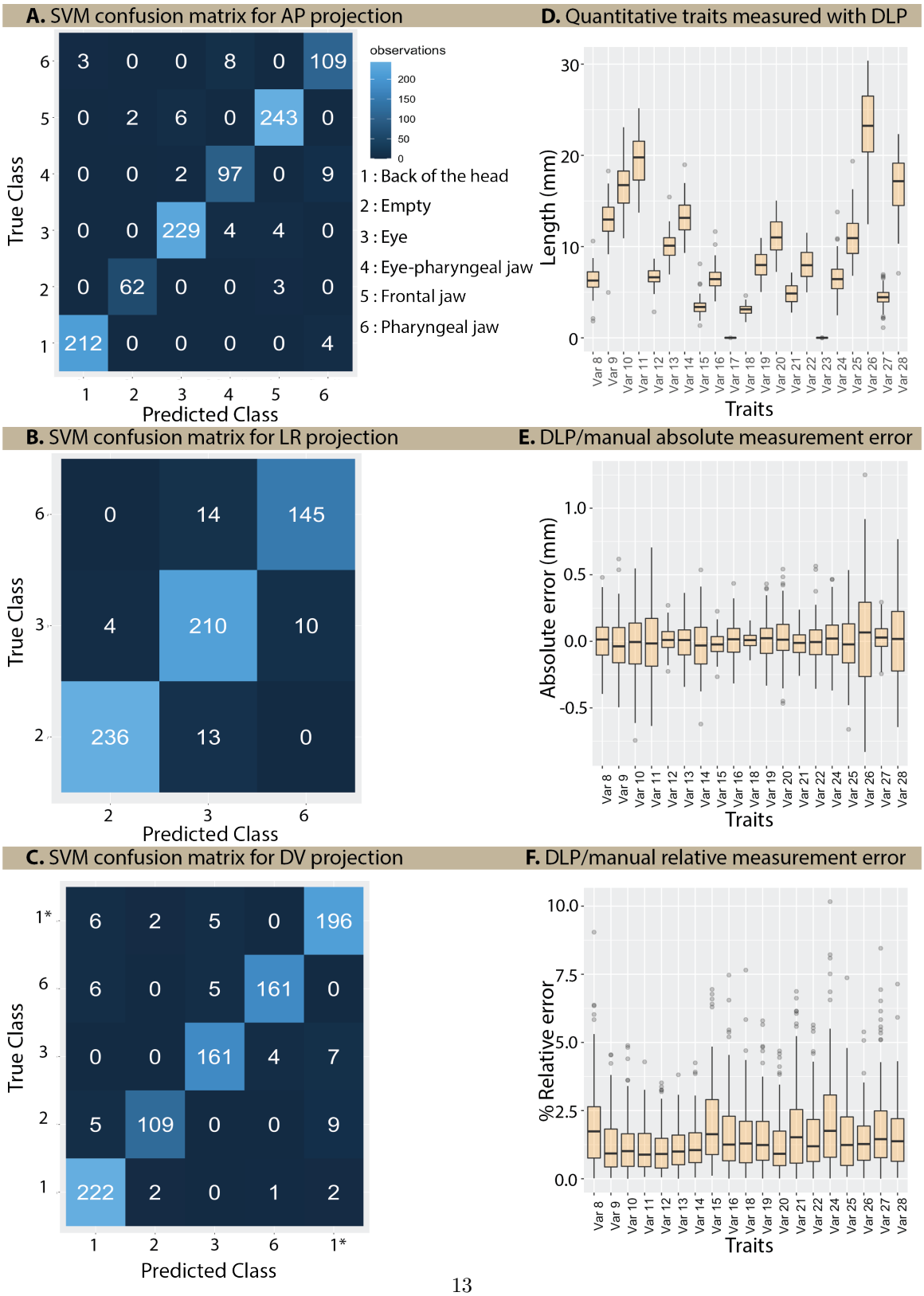

Figure 4: (A,B,C) confusion matrices for classification of slices by SVMs for each of the AP, LR and DV axes, evaluated on held out labelled data using a 50% posterior probability threshold. These give overall SVM classification accuracies of 94%, 92% and 89% for the three axes respectively. (D) Boxplot showing distributions for 19 traits and 2 controls that were measured using the DLP. The error in measurement was assessed by comparing the DLP measurements with measurements that were conducted manually using the Slicer interactive software (Kikinis et al., 2014). (E) and (F) show these error values both in absolute mm and relative to the size of the respective measurement.

relative differences vary between traits (Figure Supplementary 1) and indeed PC1 is correlated to overall size as measured by the head length (Var11) with correlation coefficient of 0.49.

The primary axis from a PCA of genome-wide variation using 926,575 SNPs with a minimum MAF of 5% describes genetic differences between species, while PC2 explains differences mainly between Thumbi West and the Otter Point populations of *M. zebroides*. Individuals of both species form distinct clusters based on locality along PCs 1 and 2, indicating clear genetic population structure according to species and geography. We controlled for spurious genetic associations arising from background genetic population structure by including the genetic PCs (1-7) that capture this population substructure as covariates in the GWAS analysis, also including Var11 (head length) as a covariate to control the systematic differences described in the previous paragraph.

As a control, we first performed a GWAS on the two technical measurement error estimates from the morphometric analysis (Var17 and Var23), for which there should be no genetic association. Indeed, neither of these genomic scans identified any significant loci. Association scans for seven of the remaining 18 traits (Var11 was used as a covariate and not tested) yielded nominally significant genetic associations across a total of 33 loci (Table 2, see Figure 5 for examples of significant associations with Posttympanic bone length (panel C) and pharyngeal jaw length (panel D)). There is substantial overlap in significant SNPs across different traits, with 5 : 10367788 associated with six traits. This is perhaps not surprising since the traits are not independent. Inclusion of the genetic PCs and head-length covariate effectively controlled for p-value inflation (Figures Supplementary 2 - Supplementary 6), offering support for the authenticity of the identified associations.

4 Discussion

The DLP package was developed to enable fast and accurate morphometric phenotyping that is almost entirety automated, and therefore is capable of scaling to large sample sizes. We have shown that it is accurate when compared to time-consuming and intrinsically more subjective manual approaches. It is also robust in that only three out of 118 samples had to be rejected as outliers. It requires relatively small amounts of training data and is fast to train and apply to new samples.

Performing morphological analysis through the classification of slices in micro-CT data is significantly easier and more accessible to implement compared to existing land-marking based techniques in three dimensions (Webster and Sheets, 2010; Gunz and Mitteroecker, 2013; Burress, 2016; Pauers et al., 2018; Hendrick et al., 2019) since the slice-based labelling process for training only takes about an hour to perform for all three projections for each 3D feature. The models trained using these labelled images can then be applied to an entire micro-CT dataset, and images from each sample do not need to be manually processed or land-marked before phenotyping.

To showcase how these measurements can be used, we trained the DLP network to investigate the morphological variation that is observed both within and between two closely related cichlid fish species: *Maylandia zebroides* and *Cynotilapia zebroides*. These measurements were then used in a GWAS for 18 of the traits in order to identify candidate genetic loci involved in determining craniofacial morphology. No significant results were found using two sets of control traits, whereas for seven skeletal traits we found statistically significant genetic associations at 33 loci. These loci are promising, but particularly in light of the small sample size for GWAS should be regarded as candidate loci for further investigation by replication with additional samples and functional exploration, for example through genetic editing or gene expression experiments, rather than proven biological associations.

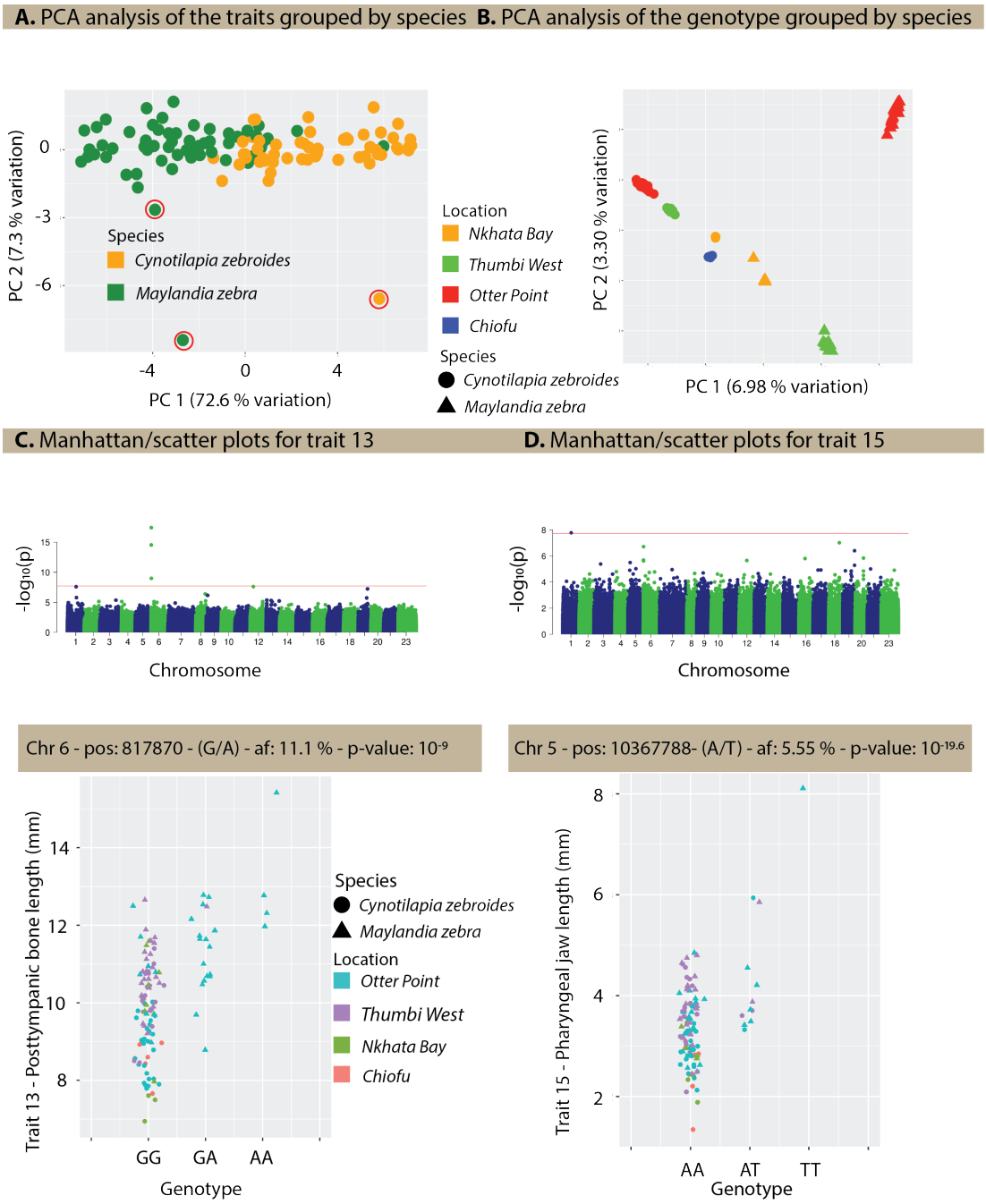

Figure 5: (A) The first two principal components of the quantitative traits measured by the DLP with points labelled according to species. The individuals shown in red are the outliers that were excluded from the GWAS. (B) The first two principal components from a PCA of genome-wide SNP variation (see section 2.1). (C) Manhattan and phenotype-genotype scatter plots showing correlations for traits 13 (Posttympanic bone length) for the genome-wide scans of association between SNPs and trait 13. In the Manhattan plots each point indicates the p-value of a SNP, with alternating blue and green colours denoting SNPs on different chromosomes. The red, horizontal line shows the 5% genome-wide significance level after a Bonferroni adjustment for multiple testing. (D) Manhattan and phenotype-genotype scatter plots showing correlations for traits 15 (Posttympanic bone length) for the genome-wide scans of association between SNPs and trait 13 (Pharyngeal jaw length).

Table 2: Statistically significant loci found for traits 8,9,12,13,14,15 and 16 with the highlighted SNPs corresponding to the loci pictured in the phenotype-genotype plots of Figure 5 (C) and (D)

Chr	Position	Frequency	Major	Minor	Traits	$-\log_{10} p$ values
1	28508896	9.9%	C	A	8,14	(157,154)
1	28508919	10.0%	T	G	15	27
2	12359064	27.6%	A	G	9,16	(8,8)
3	7173111	5.5%	C	T	12	45
3	9058204	11.7%	A	T	8,14	(12,12)
3	14695098	6.5%	A	G	9,16	(31,31)
3	48468960	5.5%	C	T	9,16	(75,76)
5	10367788	5.5%	A	T	8,9,13,14,15,16	(17,13,13,16,20,13)
5	35205044	44.5%	T	A	12	13
6	709539	10.2%	C	T	8,13,14	(7,17,7)
6	817870	11.1%	G	A	13	9
6	5698280	5.4%	C	T	9,16	(98,100)
7	8753577	7.8%	C	T	12	11
7	18398467	5.9%	T	C	8,14	(31,31)
7	25825327	7.2%	C	T	8,13,14,15	(86,200,85,9)
7	67050617	6.7%	G	A	12	24
9	22036037	5.9%	G	A	8,13,14,15	(63,229,62,19)
10	33712863	5.3%	A	G	12	40
11	29806372	5.4%	T	C	9,16	(273,284)
13	5331913	9.6%	T	A	8,13,14	(9,25,9)
14	8941641	10.7%	G	A	9,16	(9,9)
14	15181937	7.3%	T	C	8,13,14,15	(27,33,27,12)
15	25934106	23.0%	A	G	8,14	(11,11)
15	27258227	5.6%	A	G	9,16	(31,31)
15	29560314	5.5%	G	A	9,16	(210,217)
17	7685438	5.4%	C	A	12	15
19	5701806	5.9%	C	A	9,16	(48,48)
20	21728641	10.7%	A	G	9,16	(8,8)
22	2316979	5.4%	G	A	9,16	(27,27)
23	103914	5.3%	C	A	8,13,14	(17,11,17)
23	16225831	9.8%	T	C	15	78
23	31412637	5.5%	G	A	15	76

We have shown that morphological measurement of intricate skeletal elements can be carried using deep convolutional neural networks, and that overfitting can be avoided if transfer learning approaches are used, specifically based on the powerful 2D image classification network *Resnet50*. Beyond genetic studies, the DLP package can potentially be used for other quantitative skeletal analysis applications using micro-CT data. Although we only demonstrate its use here on the heads of fish, there was no specific knowledge of fish built into the package, and so we infer that it would be

applicable to other vertebrate skeletal systems, or, more generally, to other 3D datasets comprised of 2D image stacks in which 3D elements can be observed.

Acknowledgement

For assistance in sample collection we would like to acknowledge the Government of Malawi, additional members of the Cambridge Malawi cichlid field expeditions in 2016 and 2017: Milan Malinsky, Alix Tyer, George Turner, Mingliu Du, Mexford Mpulungu, Hannes Svardal, Gregoire Vernaz, Karl Svardal, and Emília Santos; and the local guides and fishermen who assisted in the collection. CT-scanning was performed by Keturah Smithson at the Cambridge Biotomography Centre, and DNA sequencing was carried out by the Scientific Operations core at the Wellcome Sanger Institute. Initial sequence data processing was performed by Hubert Denise. We would like to acknowledge the comments and feedback of Luka Moritz Blumer and Regev Schweiger on the genetics association study. This work was funded through grants from the Wellcome Trust (WT206194 to the Sanger Institute, WT207492 to R.D.) and a studentship from Trinity Hall (to A.K.). For the purpose of open access, the author has applied a CC BY public copyright licence to any Author Accepted Manuscript version arising from this submission.

Conflict of Interest Statement

The authors declare no conflict of interests.

Author Contributions

RD conceived the research problem and contributed to the method design and analysis; RZ and BR collected samples; AK developed and implemented the compression, pre-processing, and the phenotyping pipeline; TL and AK performed the GWAS; AK processed the results and led the writing of this manuscript. All authors contributed critically to the drafts and gave final approval for publication.

Inclusion statement: This study arises from a long term collaboration between scientists in Malawi and the United Kingdom. The specific methods described in this paper were developed during a masters' project in the UK for application to data from samples collected in Malawi. Ongoing research under an Access and Benefit Sharing agreement will involve local research students and scientists in experimental design and analysis.

Code and Data Availability

The code can be found on the following GitHub page:

`https://github.com/armankarshenas/DLP`

We have deposited an example dataset in dryad with doi: https://doi.org/10.6078/D1M13B containing both the primary CT data for a sample (as an Anterior-Posterior image stack), and the denoised and compressed image stacks in all three orientations (AP, Left-Right, Dorsal-Ventral) which are generated by the initial pre-processing steps described in the paper.

References

John Ashburner. A fast diffeomorphic image registration algorithm. *NeuroImage*, 38(1):95 – 113, 2007. ISSN 1053-8119. doi: https://doi.org/10.1016/j.neuroimage.2007.07.007. URL http://www.sciencedirect.com/science/article/pii/S1053811907005848.

John Ashburner and Karl J. Friston. Voxel-based morphometry—the methods. *NeuroImage*, 11 (6):805 – 821, 2000. ISSN 1053-8119. doi: https://doi.org/10.1006/nimg.2000.0582. URL http://www.sciencedirect.com/science/article/pii/S1053811900905822.

Bernhard E. Boser, Isabelle M. Guyon, and Vladimir N. Vapnik. A training algorithm for optimal margin classifiers. In *Proceedings of the Fifth Annual Workshop on Computational Learning Theory*, COLT '92, page 144–152, New York, NY, USA, 1992. Association for Computing Machinery. ISBN 089791497X. doi: 10.1145/130385.130401. URL https://doi.org/10.1145/130385.130401.

Annalisa Buniello, Jacqueline A L MacArthur, Maria Cerezo, Laura W Harris, James Hayhurst, Cinzia Malangone, Aoife McMahon, Joannella Morales, Edward Mountjoy, Elliot Sollis, Daniel Suveges, Olga Vrousgou, Patricia L Whetzel, Ridwan Amode, Jose A Guillen, Harpreet S Riat, Stephen J Trevanion, Peggy Hall, Heather Junkins, Paul Flicek, Tony Burdett, Lucia A Hindorff, Fiona Cunningham, and Helen Parkinson. The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. *Nucleic Acids Research*, 47(D1):D1005–D1012, 11 2018. ISSN 0305-1048. doi: 10.1093/nar/gky1120. URL https://doi.org/10.1093/nar/gky1120.

Edward D. Burress. Ecological diversification associated with the pharyngeal jaw diversity of neotropical cichlid fishes. *Journal of Animal Ecology*, 85(1):302–313, 2016. doi: https://doi.org/10.1111/1365-2656.12457. URL https://besjournals.onlinelibrary.wiley.com/doi/abs/10.1111/1365-2656.12457.

Paul R. Burton, David G. Clayton, Lon R. Cardon, Nick Craddock, Panos Deloukas, Audrey Duncanson, Dominic P. Kwiatkowski, Mark I. McCarthy, Willem H. Ouwehand, Nilesh J. Samani, John A. Todd, Peter Donnelly, Jeffrey C. Barrett, Dan Davison, Doug Easton, David Evans, Hin-Tak Leung, Jonathan L. Marchini, Andrew P. Morris, Chris C. A. Spencer, Martin D. Tobin, Antony P. Attwood, James P. Boorman, Barbara Cant, Ursula Everson, Judith M. Hussey, Jennifer D. Jolley, Alexandra S. Knight, Kerstin Koch, Elizabeth Meech, Sarah Nutland, Christopher V. Prowse, Helen E. Stevens, Niall C. Taylor, Graham R. Walters, Neil M. Walker, Nicholas A. Watkins, Thilo Winzer, Richard W. Jones, Wendy L. McArdle, Susan M. Ring, David P. Strachan, Marcus Pembrey, Gerome Breen, David St Clair, Sian Caesar, Katherine Gordon-Smith, Lisa Jones, Christine Fraser, Elaine K. Green, Detelina Grozeva, Marian L. Hamshere, Peter A. Holmans, Ian R. Jones, George Kirov, Valentina Moskvina, Ivan Nikolov, Michael C. O'Donovan, Michael J. Owen, David A. Collier, Amanda Elkin, Anne Farmer, Richard Williamson, Peter McGuffin, Allan H. Young, I. Nicol Ferrier, Stephen G. Ball, Anthony J. Balmforth, Jennifer H. Barrett, D. Timothy Bishop, Mark M. Iles, Azhar Maqbool, Nadira Yuldasheva, Alistair S. Hall, Peter S. Braund, Richard J. Dixon, Massimo Mangino, Suzanne Stevens, John R. Thompson, Francesca Bredin, Mark Tremelling, Miles Parkes, Hazel Drummond, Charles W. Lees, Elaine R. Nimmo, Jack Satsangi, Sheila A. Fisher, Alastair Forbes, Cathryn M. Lewis, Clive M. Onnie, Natalie J. Prescott, Jeremy Sanderson, Christopher G. Mathew, Jamie Barbour, M. Khalid Mohiuddin, Catherine E. Todhunter, John C. Mansfield, Tariq Ahmad, Fraser R. Cummings, Derek P.

Jewell, John Webster, Morris J. Brown, G. Mark Lathrop, John Connell, Anna Dominiczak, Carolina A. Braga Marcano, Beverley Burke, Richard Dobson, Johannie Gungadoo, Kate L. Lee, Patricia B. Munroe, Stephen J. Newhouse, Abiodun Onipinla, Chris Wallace, Mingzhan Xue, Mark Caulfield, Martin Farrall, Anne Barton, , The Biologics in RA Genetics Genomics (BRAGGS), Ian N. Bruce, Hannah Donovan, Steve Eyre, Paul D. Gilbert, Samantha L. Hider, Anne M. Hinks, Sally L. John, Catherine Potter, Alan J. Silman, Deborah P. M. Symmons, Wendy Thomson, Jane Worthington, David B. Dunger, Barry Widmer, Timothy M. Frayling, Rachel M. Freathy, Hana Lango, John R. B. Perry, Beverley M. Shields, Michael N. Weedon, Andrew T. Hattersley, Graham A. Hitman, Mark Walker, Kate S. Elliott, Christopher J. Groves, Cecilia M. Lindgren, Nigel W. Rayner, Nicholas J. Timpson, Eleftheria Zeggini, Melanie Newport, Giorgio Sirugo, Emily Lyons, Fredrik Vannberg, Adrian V. S. Hill, Linda A. Bradbury, Claire Farrar, Jennifer J. Pointon, Paul Wordsworth, Matthew A. Brown, Jayne A. Franklyn, Joanne M. Heward, Matthew J. Simmonds, Stephen C. L. Gough, Sheila Seal, Breast Cancer Susceptibility Collaboration (UK), Michael R. Stratton, Nazneen Rahman, Maria Ban, An Goris, Stephen J. Sawcer, Alastair Compston, David Conway, Muminatou Jallow, Kirk A. Rockett, Suzannah J. Bumpstead, Amy Chaney, Kate Downes, Mohammed J. R. Ghori, Rhian Gwilliam, Sarah E. Hunt, Michael Inouye, Andrew Keniry, Emma King, Ralph McGinnis, Simon Potter, Rathi Ravindrarajah, Pamela Whittaker, Claire Widden, David Withers, Niall J. Cardin, Teresa Ferreira, Joanne Pereira-Gale, Ingileif B. Hallgrimsdóttir, Bryan N. Howie, Zhan Su, Yik Ying Teo, Damjan Vukcevic, David Bentley, Alistair Compston, The Wellcome Trust Case Control Consortium, Management Committee, Data, Analysis Committee, UK Blood Services, University of Cambridge Controls, 1958 Birth Cohort Controls, Bipolar Disorder, Coronary Artery Disease, Crohn's Disease, Hypertension, Rheumatoid Arthritis, Type 1 Diabetes, Type 2 Diabetes, Tuberculosis, Ankylosing Spondylitis, Autoimmune Thyroid Disease, Breast Cancer, Multiple Sclerosis, Gambian Controls, Data QC DNA, Genotyping, Informatics, Statistics, and Primary Investigators. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature*, 447(7145):661–678, Jun 2007. ISSN 1476-4687. doi: 10.1038/nature05911. URL https://doi.org/10.1038/nature05911.

Stephen Butterworth et al. On the theory of filter amplifiers. *Wireless Engineer*, 7(6):536–541, 1930.

Liang-Chieh Chen, Yukun Zhu, George Papandreou, Florian Schroff, and Hartwig Adam. Encoder-decoder with atrous separable convolution for semantic image segmentation. *CoRR*, abs/1802.02611, 2018. URL http://arxiv.org/abs/1802.02611.

Jerry A. Coyne. Genetics and speciation. *Nature*, 355(6360):511–515, Feb 1992. ISSN 1476-4687. doi: 10.1038/355511a0. URL https://doi.org/10.1038/355511a0.

Cynthia Ebinger. Continental break-up: The East African perspective. *Astronomy & Geophysics*, 46(2):2.16–2.21, 04 2005. ISSN 1366-8781. doi: 10.1111/j.1468-4004.2005.46216.x. URL https://doi.org/10.1111/j.1468-4004.2005.46216.x.

Santiago Fernández, Alex Graves, and Jürgen Schmidhuber. An application of recurrent neural networks to discriminative keyword spotting. In *Proceedings of the 17th International Conference on Artificial Neural Networks*, ICANN'07, page 220–229, Berlin, Heidelberg, 2007. Springer-Verlag. ISBN 3540746935.

Anne Frary, T. Clint Nesbitt, Amy Frary, Silvana Grandillo, Esther van der Knaap, Bin Cong, Jiping Liu, Jaroslaw Meller, Ron Elber, Kevin B. Alpert, and Steven D. Tanksley. fw2.2: A quantitative

trait locus key to the evolution of tomato fruit size. *Science*, 289(5476):85–88, 2000. ISSN 0036-8075. doi: 10.1126/science.289.5476.85. URL https://science.sciencemag.org/content/289/5476/85.

Karl Friston, John Ashburner, Stefan Kiebel, Thomas Nichols, and William Penny. *CHAPTER 1 - A short history of SPM*. Academic Press, London, 2007. ISBN 978-0-12-372560-8. doi: https://doi.org/10.1016/B978-012372560-8/50001-2. URL https://www.sciencedirect.com/science/article/pii/B9780123725608500012.

Matteo Fumagalli, Filipe G. Vieira, Tyler Linderoth, and Rasmus Nielsen. ngsTools: methods for population genetics analyses from next-generation sequencing data. *Bioinformatics*, 30(10):1486–1487, 01 2014. ISSN 1367-4803. doi: 10.1093/bioinformatics/btu041. URL https://doi.org/10.1093/bioinformatics/btu041.

Sorin Mihai Grigorescu, Bogdan Trasnea, Tiberiu T. Cocias, and Gigel Macesanu. A survey of deep learning techniques for autonomous driving. *CoRR*, abs/1910.07738, 2019. URL http://arxiv.org/abs/1910.07738.

Philipp Gunz and Philipp Mitteroecker. Semilandmarks: a method for quantifying curves and surfaces. *Hystrix, the Italian Journal of Mammalogy*, 24(1):103–109, 2013. ISSN 0394-1914. doi: 10.4404/hystrix-24.1-6292. URL https://doi.org/10.4404/hystrix-24.1-6292.

Mohammad Havaei, Axel Davy, David Warde-Farley, Antoine Biard, Aaron C. Courville, Yoshua Bengio, Chris Pal, Pierre-Marc Jodoin, and Hugo Larochelle. Brain tumor segmentation with deep neural networks. *CoRR*, abs/1505.03540, 2015. URL http://arxiv.org/abs/1505.03540.

Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition, 2015a.

Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. *CoRR*, abs/1512.03385, 2015b. URL http://arxiv.org/abs/1512.03385.

Yajun He, Daoming Wu, Dayong Wei, Ying Fu, Yixin Cui, Hongli Dong, Chuandong Tan, and Wei Qian. GWAS, QTL mapping and gene expression analyses in Brassica napus reveal genetic control of branching morphogenesis. *Scientific Reports*, 7(1):15971, Nov 2017. ISSN 2045-2322. doi: 10.1038/s41598-017-15976-4. URL https://doi.org/10.1038/s41598-017-15976-4.

Laura A. Hendrick, Grace A. Carter, Erin H. Hilbrands, Brian P. Heubel, Thomas F. Schilling, and Pierre Le Pabic. Bar, stripe and spot development in sand-dwelling cichlid from Lake Malawi. *EvoDevo*, 10(1):18, Aug 2019. ISSN 2041-9139. doi: 10.1186/s13227-019-0132-7. URL https://doi.org/10.1186/s13227-019-0132-7.

Park Ji Wan Hong Eun Pyo. Sample size and statistical power calculation in genetic association studies. *Genomics Inform*, 10(2):117–122, 2012. doi: 10.5808/GI.2012.10.2.117. URL http://genominfo.org/journal/view.php?number=5.

G. N. Hounsfield. Computerized transverse axial scanning (tomography): Part 1. description of system. *The British Journal of Radiology*, 46(552):1016–1022, 1973. doi: 10.1259/0007-1285-46-552-1016. URL https://doi.org/10.1259/0007-1285-46-552-1016. PMID: 4757352.

Godfrey N. Hounsfield. Computed medical imaging. *Science*, 210(4465):22–28, 1980. doi: 10.1126/science.6997993. URL https://www.science.org/doi/abs/10.1126/science.6997993.

David A. Huffman. A method for the construction of minimum-redundancy codes. *Proceedings of the IRE*, 40(9):1098–1101, 1952. doi: 10.1109/JRPROC.1952.273898.

Thomas C. Johnson, Christopher A. Scholz, Michael R. Talbot, Kerry Kelts, R. D. Ricketts, Gideon Ngobi, Kristina Beuning, Immacculate Ssemmanda, and J. W. McGill. Late pleistocene desiccation of Lake Victoria and rapid evolution of cichlid fishes. *Science*, 273(5278):1091–1093, 1996. ISSN 0036-8075. doi: 10.1126/science.273.5278.1091. URL https://science.sciencemag.org/content/273/5278/1091.

Artem Khmelinskii, Harald C. Groen, Martin Baiker, Marion de Jong, and Boudewijn P. F. Lelieveldt. Segmentation and visual analysis of whole-body mouse skeleton microspect. *PLOS ONE*, 7(11):1–10, 11 2012. doi: 10.1371/journal.pone.0048976. URL https://doi.org/10.1371/journal.pone.0048976.

Ron Kikinis, Steve D. Pieper, and Kirby G. Vosburgh. *3D Slicer: A Platform for Subject-Specific Image Analysis, Visualization, and Clinical Support*, pages 277–289. Springer New York, New York, NY, 2014. ISBN 978-1-4614-7657-3. doi: 10.1007/978-1-4614-7657-3_19. URL https://doi.org/10.1007/978-1-4614-7657-3_19.

Thorfinn Sand Korneliussen, Anders Albrechtsen, and Rasmus Nielsen. ANGSD: Analysis of Next Generation Sequencing Data. *BMC Bioinformatics*, 15(1):356, Nov 2014. ISSN 1471-2105. doi: 10.1186/s12859-014-0356-4. URL https://doi.org/10.1186/s12859-014-0356-4.

Irv Kornfield and Peter F. Smith. African cichlid fishes: Model systems for evolutionary biology. *Annual Review of Ecology and Systematics*, 31(1):163–196, 2000. doi: 10.1146/annurev.ecolsys.31.1.163. URL https://doi.org/10.1146/annurev.ecolsys.31.1.163.

Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. In F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger, editors, *Advances in Neural Information Processing Systems*, volume 25, pages 1097–1105. Curran Associates, Inc., 2012. URL https://proceedings.neurips.cc/paper/2012/file/c399862d3b9d6b76c8436e924a68c45b-Paper.pdf.

Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. Deep learning. *Nature*, 521(7553):436–444, May 2015. ISSN 1476-4687. doi: 10.1038/nature14539. URL https://doi.org/10.1038/nature14539.

Yan Li and Liming Xia. Coronavirus disease 2019 (covid-19): Role of chest ct in diagnosis and management. *American Journal of Roentgenology*, 214(6):1280–1286, 2020. doi: 10.2214/AJR.20.22954. URL https://doi.org/10.2214/AJR.20.22954. PMID: 32130038.

Tsung-Yi Lin, Michael Maire, Serge J. Belongie, Lubomir D. Bourdev, Ross B. Girshick, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C. Lawrence Zitnick. Microsoft COCO: common objects in context. *CoRR*, abs/1405.0312, 2014. URL http://arxiv.org/abs/1405.0312.

Milan Malinsky, Hannes Svardal, Alexandra M. Tyers, Eric A. Miska, Martin J. Genner, George F. Turner, and Richard Durbin. Whole-genome sequences of malawi cichlid reveal multiple radiations interconnected by gene flow. *Nature Ecology & Evolution*, 2(12):1940–1955, Dec 2018. ISSN 2397-334X. doi: 10.1038/s41559-018-0717-x. URL https://doi.org/10.1038/s41559-018-0717-x.

Ichiro Matsuda, Nau Ozaki, Yuji Umezu, and Susumu Itoh. Lossless coding using variable block-size adaptive prediction optimized for each image. In *2005 13th European Signal Processing Conference*, pages 1–4. IEEE, 2005.

Matthew D. McGee, Samuel R. Borstein, Joana I. Meier, David A. Marques, Salome Mwaiko, Anthony Taabu, Mary A. Kishe, Brian O'Meara, Rémy Bruggmann, Laurent Excoffier, and Ole Seehausen. The ecological and genomic basis of explosive adaptive radiation. *Nature*, 586(7827): 75–79, Oct 2020. ISSN 1476-4687. doi: 10.1038/s41586-020-2652-7. URL https://doi.org/10.1038/s41586-020-2652-7.

Aaron McKenna, Matthew Hanna, Eric Banks, Andrey Sivachenko, Kristian Cibulskis, Andrew Kernytsky, Kiran Garimella, David Altshuler, Stacey Gabriel, Mark Daly, and Mark A. DePristo. The genome analysis toolkit: A mapreduce framework for analyzing next-generation dna sequencing data. *Genome Research*, 20(9):1297–1303, 2010. doi: 10.1101/gr.107524.110. URL http://genome.cshlp.org/content/20/9/1297.abstract.

Jishnu Mukhoti and Yarin Gal. Evaluating bayesian deep learning methods for semantic segmentation. *CoRR*, abs/1811.12709, 2018. URL http://arxiv.org/abs/1811.12709.

Patrik Nosil, Jeffrey L. Feder, and Zachariah Gompert. How many genetic changes create new species? *Science*, 371(6531):777–779, 2021. ISSN 0036-8075. doi: 10.1126/science.abf6671. URL https://science.sciencemag.org/content/371/6531/777.

Sinno Jialin Pan and Qiang Yang. A survey on transfer learning. *IEEE Transactions on Knowledge and Data Engineering*, 22(10):1345–1359, 2010. doi: 10.1109/TKDE.2009.191.

Michael J. Pauers, Kelsey R. Fox, Robert A. Hall, and Kesha Patel. Selection, hybridization, and the evolution of morphology in the Lake Malaŵi endemic cichlid of the genus labeotropheus. *Scientific Reports*, 8(1):15842, Oct 2018. ISSN 2045-2322. doi: 10.1038/s41598-018-34135-x. URL https://doi.org/10.1038/s41598-018-34135-x.

Jie Peng, Shuai Kang, Zhengyuan Ning, Hangxia Deng, Jingxian Shen, Yikai Xu, Jing Zhang, Wei Zhao, Xinling Li, Wuxing Gong, Jinhua Huang, and Li Liu. Residual convolutional neural network for predicting response of transarterial chemoembolization in hepatocellular carcinoma from ct imaging. *European Radiology*, 30(1):413–424, Jan 2020. ISSN 1432-1084. doi: 10.1007/s00330-019-06318-1. URL https://doi.org/10.1007/s00330-019-06318-1.

Joseph Redmon and Ali Farhadi. YOLO9000: better, faster, stronger. *CoRR*, abs/1612.08242, 2016. URL http://arxiv.org/abs/1612.08242.

Joseph Redmon and Ali Farhadi. Yolov3: An incremental improvement. *CoRR*, abs/1804.02767, 2018. URL http://arxiv.org/abs/1804.02767.

Joseph Redmon, Santosh Kumar Divvala, Ross B. Girshick, and Ali Farhadi. You only look once: Unified, real-time object detection. *CoRR*, abs/1506.02640, 2015. URL http://arxiv.org/abs/1506.02640.

Derek A. Roff and Timothy A. Mousseau. Quantitative genetics and fitness: lessons from drosophila. *Heredity*, 58(1):103–118, Feb 1987. ISSN 1365-2540. doi: 10.1038/hdy.1987.15. URL https://doi.org/10.1038/hdy.1987.15.

Fabrizia Ronco, Michael Matschiner, Astrid Böhne, Anna Boila, Heinz H. Büscher, Athimed El Taher, Adrian Indermaur, Milan Malinsky, Virginie Ricci, Ansgar Kahmen, Sissel Jentoft, and Walter Salzburger. Drivers and dynamics of a massive adaptive radiation in cichlid fishes. *Nature*, 589(7840):76–81, Jan 2021. ISSN 1476-4687. doi: 10.1038/s41586-020-2930-4. URL https://doi.org/10.1038/s41586-020-2930-4.

Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, Alexander C. Berg, and Li Fei-Fei. ImageNet Large Scale Visual Recognition Challenge. *International Journal of Computer Vision (IJCV)*, 115 (3):211–252, 2015. doi: 10.1007/s11263-015-0816-y.

Dolph Schluter. Evidence for ecological speciation and its alternative. *Science*, 323(5915):737–741, 2009. ISSN 0036-8075. doi: 10.1126/science.1160006. URL https://science.sciencemag.org/content/323/5915/737.

L. Skotte, T. S. Korneliussen, and A. Albrechtsen. Association testing for next-generation sequencing data using score statistics. *Genet Epidemiol*, 36(5):430–437, Jul 2012.

Marcus Soliai, Agnes S. Sundaresan, Andréanne Morin, Annemarie G. Hirsch, Catherine Stanhope, Jordan Kuiper, Brian S. Schwartz, Carole Ober, and Jayant M. Pinto. Two-stage genome-wide association study of chronic rhinosinusitis and disease subphenotypes highlights mucosal immunity contributing to risk. *International Forum of Allergy & Rhinology*, 11(4):814–817, 2021. doi: https://doi.org/10.1002/alr.22731. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/alr.22731.

Chris C. A. Spencer, Zhan Su, Peter Donnelly, and Jonathan Marchini. Designing genome-wide association studies: Sample size, power, imputation, and the choice of genotyping chip. *PLOS Genetics*, 5(5):1–13, 05 2009. doi: 10.1371/journal.pgen.1000477. URL https://doi.org/10.1371/journal.pgen.1000477.

Sebastian Steger, Matthias Kirschner, and Stefan Wesarg. Articulated atlas for segmentation of the skeleton from head amp; neck ct datasets. In *2012 9th IEEE International Symposium on Biomedical Imaging (ISBI)*, pages 1256–1259, 2012. doi: 10.1109/ISBI.2012.6235790.

Hannes Svardal, Fu Xiang Quah, Milan Malinsky, Benjamin P Ngatunga, Eric A Miska, Walter Salzburger, Martin J Genner, George F Turner, and Richard Durbin. Ancestral Hybridization Facilitated Species Diversification in the Lake Malawi Cichlid Fish Adaptive Radiation. *Molecular Biology and Evolution*, 37(4):1100–1113, 12 2019. ISSN 0737-4038. doi: 10.1093/molbev/msz294. URL https://doi.org/10.1093/molbev/msz294.

Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott E. Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. Going deeper with convolutions. *CoRR*, abs/1409.4842, 2014. URL http://arxiv.org/abs/1409.4842.

Peter M. Visscher, William G. Hill, and Naomi R. Wray. Heritability in the genomics era —concepts and misconceptions. *Nature Reviews Genetics*, 9(4):255–266, 2008. doi: 10.1038/nrg2322. URL https://doi.org/10.1038/nrg2322.

J. G. Wallace, S. J. Larsson, and E. S. Buckler. Entering the second century of maize quantitative genetics. *Heredity*, 112(1):30–38, Jan 2014. ISSN 1365-2540. doi: 10.1038/hdy.2013.6. URL https://doi.org/10.1038/hdy.2013.6.

Mark Webster and H. David Sheets. A practical introduction to landmark-based geometric morphometrics. *The Paleontological Society Papers*, 16:163–188, 2010. doi: 10.1017/S1089332600001868.

Karl Weiss, Taghi M. Khoshgoftaar, and DingDing Wang. A survey of transfer learning. *Journal of Big Data*, 3(1):9, May 2016. ISSN 2196-1115. doi: 10.1186/s40537-016-0043-6. URL https://doi.org/10.1186/s40537-016-0043-6.

Philip J. Withers, Charles Bouman, Simone Carmignato, Veerle Cnudde, David Grimaldi, Charlotte K. Hagen, Eric Maire, Marena Manley, Anton Du Plessis, and Stuart R. Stock. X-ray computed tomography. *Nature Reviews Methods Primers*, 1(1):18, Feb 2021. ISSN 2662-8449. doi: 10.1038/s43586-021-00015-4. URL https://doi.org/10.1038/s43586-021-00015-4.

Andrew R. Wood, Tonu Esko, Jian Yang, Sailaja Vedantam, Tune H. Pers, Stefan Gustafsson, Audrey Y. Chu, Karol Estrada, Jian'an Luan, Zoltán Kutalik, Najaf Amin, Martin L. Buchkovich, Damien C. Croteau-Chonka, Felix R. Day, Yanan Duan, Tove Fall, Rudolf Fehrmann, Teresa Ferreira, Anne U. Jackson, Juha Karjalainen, Ken Sin Lo, Adam E. Locke, Reedik Mägi, Evelin Mihailov, Eleonora Porcu, Joshua C. Randall, André Scherag, Anna A. E. Vinkhuyzen, Harm-Jan Westra, Thomas W. Winkler, Tsegaselassie Workalemahu, Jing Hua Zhao, Devin Absher, Eva Albrecht, Denise Anderson, Jeffrey Baron, Marian Beekman, Ayse Demirkan, Georg B. Ehret, Bjarke Feenstra, Mary F. Feitosa, Krista Fischer, Ross M. Fraser, Anuj Goel, Jian Gong, Anne E. Justice, Stavroula Kanoni, Marcus E. Kleber, Kati Kristiansson, Unhee Lim, Vaneet Lotay, Julian C. Lui, Massimo Mangino, Irene Mateo Leach, Carolina Medina-Gomez, Michael A. Nalls, Dale R. Nyholt, Cameron D. Palmer, Dorota Pasko, Sonali Pechlivanis, Inga Prokopenko, Janina S. Ried, Stephan Ripke, Dmitry Shungin, Alena Stancáková, Rona J. Strawbridge, Yun Ju Sung, Toshiko Tanaka, Alexander Teumer, Stella Trompet, Sander W. van der Laan, Jessica van Setten, Jana V. Van Vliet-Ostaptchouk, Zhaoming Wang, Loïc Yengo, Weihua Zhang, Uzma Afzal, Johan Ärnlöv, Gillian M. Arscott, Stefania Bandinelli, Amy Barrett, Claire Bellis, Amanda J. Bennett, Christian Berne, Matthias Blüher, Jennifer L. Bolton, Yvonne Böttcher, Heather A. Boyd, Marcel Bruinenberg, Brendan M. Buckley, Steven Buyske, Ida H. Caspersen, Peter S. Chines, Robert Clarke, Simone Claudi-Boehm, Matthew Cooper, E. Warwick Daw, Pim A. De Jong, Joris Deelen, Graciela Delgado, Josh C. Denny, Rosalie Dhonukshe-Rutten, Maria Dimitriou, Alex S. F. Doney, Marcus Dörr, Niina Eklund, Elodie Eury, Lasse Folkersen, Melissa E. Garcia, Frank Geller, Vilmantas Giedraitis, Alan S. Go, Harald Grallert, Tanja B. Grammer, Jürgen Gräßler, Henrik Grönberg, Lisette C. P. G. M. de Groot, Christopher J. Groves, Jeffrey Haessler, Per Hall, Toomas Haller, Goran Hallmans, Anke Hannemann, Catharina A. Hartman, Maija Hassinen, Caroline Hayward, Nancy L. Heard-Costa, Quinta Helmer, Gibran Hemani, Anjali K. Henders, Hans L. Hillege, Mark A. Hlatky, Wolfgang Hoffmann, Per Hoffmann, Oddgeir Holmen, Jeanine J. Houwing-Duistermaat, Thomas Illig, Aaron Isaacs, Alan L. James, Janina Jeff, Berit Johansen, Åsa Johansson, Jennifer Jolley, Thorhildur Juliusdottir, Juhani Junttila, Abel N. Kho, Leena Kinnunen, Norman Klopp, Thomas Kocher, Wolfgang Kratzer, Peter Lichtner, Lars Lind, Jaana Lindström, Stéphane Lobbens, Mattias Lorentzon, Yingchang Lu, Valeriya Lyssenko, Patrik K. E. Magnusson, Anubha Mahajan, Marc Maillard, Wendy L. McArdle, Colin A. McKenzie, Stela McLachlan, Paul J. McLaren, Cristina Menni, Sigrun Merger, Lili Milani, Alireza Moayyeri, Keri L. Monda, Mario A. Morken, Gabriele Müller, Martina Müller-Nurasyid, Arthur W. Musk, Narisu Narisu, Matthias Nauck, Ilja M. Nolte, Markus M. Nöthen, Laticia Oozageer, Stefan Pilz, Nigel W. Rayner, Frida Renstrom, Neil R. Robertson, Lynda M. Rose, Ronan Roussel, Serena Sanna, Hubert Scharnagl, Salome Scholtens, Fredrick R. Schumacher, Heribert Schunkert, Robert A. Scott, Joban Sehmi, Thomas Seufferlein, Jianxin Shi, Karri Silventoinen, Johannes H. Smit, Albert Vernon Smith, Joanna Smolonska, Alice V. Stanton, Kathleen Stirrups, David J. Stott, Heather M. Stringham, Johan Sundström, Morris A. Swertz, Ann-Christine Syvänen, Bamidele O. Tayo, Gudmar Thorleifsson, Jonathan P. Tyrer, Suzanne van Dijk, Natasja M. van Schoor, Nathalie van der Velde, Diana van Heemst, Floor V. A. van Oort, Sita H. Vermeulen, Niek Verweij, Judith M. Vonk, Lindsay L. Waite, Melanie Waldenberger, Roman Wennauer, Lynne R. Wilkens, Christina Willenborg, Tom Wilsgaard, Mary K. Wojczynski, Andrew Wong, Alan F. Wright, Qunyuan Zhang, Dominique Arveiler, Stephan J. L. Bakker, John Beilby, Richard N. Bergman, Sven Bergmann,

Reiner Biffar, John Blangero, Dorret I. Boomsma, Stefan R. Bornstein, Pascal Bovet, Paolo Brambilla, Morris J. Brown, Harry Campbell, Mark J. Caulfield, Aravinda Chakravarti, Rory Collins, Francis S. Collins, Dana C. Crawford, L. Adrienne Cupples, John Danesh, Ulf de Faire, Hester M. den Ruijter, Raimund Erbel, Jeanette Erdmann, Johan G. Eriksson, Martin Farrall, Ele Ferrannini, Jean Ferrières, Ian Ford, Nita G. Forouhi, Terrence Forrester, Ron T. Gansevoort, Pablo V. Gejman, Christian Gieger, Alain Golay, Omri Gottesman, Vilmundur Gudnason, Ulf Gyllensten, David W. Haas, Alistair S. Hall, Tamara B. Harris, Andrew T. Hattersley, Andrew C. Heath, Christian Hengstenberg, Andrew A. Hicks, Lucia A. Hindorff, Aroon D. Hingorani, Albert Hofman, G. Kees Hovingh, Steve E. Humphries, Steven C. Hunt, Elina Hypponen, Kevin B. Jacobs, Marjo-Riitta Jarvelin, Pekka Jousilahti, Antti M. Jula, Jaakko Kaprio, John J. P. Kastelein, Manfred Kayser, Frank Kee, Sirkka M. Keinanen-Kiukaanniemi, Lambertus A. Kiemeney, Jaspal S. Kooner, Charles Kooperberg, Seppo Koskinen, Peter Kovacs, Aldi T. Kraja, Meena Kumari, Johanna Kuusisto, Timo A. Lakka, Claudia Langenberg, Loic Le Marchand, Terho Lehtimäki, Sara Lupoli, and Pamela A. F. Madden. Defining the role of common variation in the genomic and biological architecture of adult human height. *Nature Genetics*, 46(11):1173–1186, Nov 2014. ISSN 1546-1718. doi: 10.1038/ng.3097. URL https://doi.org/10.1038/ng.3097.