G3, 2022, **12(2)**, jkab443
https://doi.org/10.1093/g3journal/jkab443
Advance Access Publication Date: 31 December 2021
Investigation

Testing the accuracy of 3D automatic landmarking via genome-wide association studies

Yoland Savriama * and Diethard Tautz *

Department Evolutionary Genetics, Max-Planck Institute for Evolutionary Biology, 24306 Plön, Germany

*Corresponding author: Max-Planck Institute for Evolutionary Biology, 24306 Plön, Germany. Email: savriama@evolbio.mpg.de; tautz@evolbio.mpg.de

Abstract

Various advances in 3D automatic phenotyping and landmark-based geometric morphometric methods have been made. While it is generally accepted that automatic landmarking compromises the capture of the biological variation, no studies have directly tested the actual impact of such landmarking approaches in analyses requiring a large number of specimens and for which the precision of phenotyping is crucial to extract an actual biological signal adequately. Here, we use a recently developed 3D atlas-based automatic landmarking method to test its accuracy in detecting QTLs associated with craniofacial development of the house mouse skull and lower jaws for a large number of specimens (circa 700) that were previously phenotyped via a semiautomatic landmarking method complemented with manual adjustment. We compare both landmarking methods with univariate and multivariate mapping of the skull and the lower jaws. We find that most significant SNPs and QTLs are not recovered based on the data derived from the automatic landmarking method. Our results thus confirm the notion that information is lost in the automated landmarking procedure although somewhat dependent on the analyzed structure. The automatic method seems to capture certain types of structures slightly better, such as lower jaws whose shape is almost entirely summarized by its outline and could be assimilated as a 2D flat object. By contrast, the more apparent 3D features exhibited by a structure such as the skull are not adequately captured by the automatic method. We conclude that using 3D atlas-based automatic landmarking methods requires careful consideration of the experimental question.

Keywords: automatic phenotyping; atlas-based segmentation; 3D landmarking; geometric morphometrics; skull; lower jaws; QTL mapping; GWAS; *Mus musculus domesticus*

Introduction

Organismal biology is now entering the so-called "Era of Big Data" (e.g. Muñoz and Price 2019) and "Phenomics" (Houle et al. 2010). From the morphological standpoint, this trend reflects the unprecedented progress with analytical devices for the 3D high-definition imaging of morphological phenotypes alongside technological and software development to collect, manage, store, and process the large amount of data produced. One key aspect that must remain central to this revolution in digital morphology is our ability to develop conceptual and methodological frameworks to extract biologically relevant information from these data and address important questions in evolutionary, functional, and developmental morphology. This is where morphometrics, the statistical analysis of shape, plays a major role.

In particular, landmark-based geometric morphometric methods have grown into a mature and powerful branch of biometrics, allowing the capture of morphologically meaningful signals from the diversity and richness of modern imaging data types. These frameworks use a mathematical description of biological structures according to geometric definitions of their size and shape extracted from Cartesian coordinates of points traditionally placed at recognizable structures across specimens (e.g. the intersection of cranial sutures or maximum of bone curvature). Such tasks are particularly highly time-consuming, tedious, and can cause an error, especially when substantial sample sizes are involved as in systematics (e.g. Frost et al. 2003) and craniofacial mapping (e.g. Pallares et al. 2015; Weiss et al. 2015; Maga et al. 2017; Pavlicev et al. 2016).

Recently, several atlas-based methods for both automatic segmentation and landmarking have been developed and hold promise for a streamlined treatment of extensive datasets (Maga et al. 2017; Percival et al. 2019; Devine et al. 2020; Porto et al. 2021). These methods use an average volume computed from all available specimens or a subset of them (or a representative specimen), henceforth called an atlas, and apply a suite of image registrations followed by deformations to segment all specimens subsequently. These transformations are reused to propagate the atlas' landmark configuration onto all segmented structures (Maga et al. 2017; Percival et al. 2019; Porto et al. 2021) with further refinement that has been suggested to reduce the known systematic difference in means and variance–covariance structure between manual and such automatic procedures (Percival et al. 2019; Porto et al. 2021).

Here, we compare the 2 approaches by running separate full genome-wide association analyses, which typically require a very

Received: August 27, 2021. **Accepted:** December 16, 2021

high number of specimens and precise phenotyping to identify genes associated with craniofacial development (e.g. Burgio et al. 2009; Pallares et al. 2015; Navarro and Maga 2016; Maga et al. 2017; Katz et al. 2020). To this aim, we reanalyzed a previously published dataset of nearly 700 skulls and lower jaws' shape and size in an outbred population of male mice acquired via the TINA semiautomatic landmarking tool completed with further manual adjustment (Pallares et al. 2015). The same specimens were processed and automatically landmarked via the Advanced Normalization Tools (ANTs) software (Tustison et al. 2021) and ANTsR (Avants et al. 2015; Tustison et al. 2015) pipelines using a single atlas computed from all of them (Maga et al. 2017; Percival et al. 2019).

In our study, the precision of automatic landmarking is assessed by whether or not (1) the same QTLs identified in Pallares et al. (2015) are also recovered, (2) the automatic pipeline detects at least the major QTLs, and (3) the QTLs discovered by the automatic method have their regions overlapping with the ones that have been initially detected in Pallares et al. (2015).

Materials and methods

Specimens and data

The specimens analyzed in the present study are the same (circa 700) Carworth Farms White (CFW) outbred mice (Parker et al. 2014) previously analyzed in Pallares et al. (2015), in which these authors used semiautomatic 3D landmark-based geometric morphometrics with additional manual adjustment and high-density genotyping to map genes involved in the craniofacial shape of skulls and lower jaws. All specimens were male and almost the same age (within a 2 weeks window). Pallares et al. (2015) did not show any significant correlation of age with shape variation, so age was not used as a covariate in Pallares et al. (2015).

The genotype data include a filtered set of 80,142 SNPs once genomic markers with a maximum probability genotype above 0.5 and minor allele frequencies below 2% have been excluded prior to the analyses. Morphometric and genomic data were the ones deposited by Pallares et al. (2015) on Dryad (http://dx.doi.org/10.5061/dryad.k543p - last accessed Dec 2021).

CT-scanning

Heads in Pallares et al. (2015) were scanned at a cubic voxel resolution of 21 μm using a μCT Scanco vivaCT 40 (Bruettisellen, Switzerland).

Landmarks

A configuration of 44 3D landmarks was used for the skull, with 34 placed as mirror images with respect to the midsagittal plane (paired landmarks) and 10 landmarks on top of it (unpaired landmarks). A set of 26 landmarks was used for the lower jaws (13 for each side). See Fig. 1 and Supplementary Table 1 for their definition (same as in Pallares et al. 2015).

Semiautomatic landmarking procedure

The 3D landmark data collection in Pallares et al. (2015) was achieved via a semiautomatic method implemented in the freely available TINA landmarking tool (Bromiley et al. 2014). First, the user computes an average landmark configuration from 10 manually landmarked mice. Second, 4 of these landmarks are manually placed onto each specimen to be used for global registration. Third, a multistage registration process using data generated in the previous step is used to propagate the appropriately transformed average reference configuration onto each specimen. The software defines a confidence threshold for each landmark by highlighting them, allowing the user for further manual adjustment whenever needed (Pallares et al. 2015, 2016).

Automatic landmarking procedure (ANTs software and ANTsR R package)

Here, we use and extend a recently published full 3D volumetric atlas-based image registration and deformation procedure to automatically segment and generate landmarks on subsequently isolated skulls and lower jaws based on a single manually landmarked average atlas computed from all mice CT scans (Avants et al. 2009, 2011; Maga et al. 2017; Tustison et al. 2021) https://github.com/muratmaga/mouse_CT_atlas (last accessed Dec. 2021).

This procedure uses state-of-the-art medical image registration and a segmentation toolkit implemented in the ANTs software (Avants et al. 2009) https://github.com/ANTsX/ANTs and in its R counterpart ANTsR (Avants et al. 2015; Tustison et al. 2015, 2021) https://github.com/ANTsX/ANTsR (last accessed Dec. 2021). First, since this pipeline is computationally demanding and typically generates a large amount of data (here 826 GBs), due to the numerous processing steps which output new files each time and especially with such a large sample size (circa 700 CT scans), we reduced the resolution of all scans prior to the analysis. For this purpose, DICOM files of each CT scan were imported in FIJI (Schindelin et al. 2012), converted to 8 bits data, downsampled to an isotropic 0.14-mm resolution (same as in Maga et al. 2017) using the Scale option (scaling factor 0.15 with the option "Preserve physical image dimensions" activated and with linear interpolation) in the TransformJ plugin (Meijering et al. 2001), exported as single .nii files using the FIJI NIFTI plugin (Williams 2005), and subsequently compressed using gunzip compression as .nii.gz files. These downsampled files accounted for about 1 Mb of file size each so that they could be processed faster in the ANTsR pipeline, while still allowing for the extraction of the main patterns of morphological variation (Maga et al. 2017; Porto et al. 2021).

Second, ANTs were used to compute a single average atlas from all downsampled files generated in the previous step using the script "antsMultivariateTemplateConstruction2.sh" (Avants et al. 2011) https://github.com/ntustison/TemplateBuilding Example (last accessed Dec. 2021). Since specimens were scanned in a standard orientation and consequently were already spatially close to each other, no preliminary registration prior to the template computation was needed to correct for differences in pitch, roll, and yaw, and the default registration steps already included in the ANTs template building script were sufficient. Eight rounds of iteration were satisfactory to generate a template detailed enough to segment the structures of interest for this study, particularly since it was generated from nearly 700 specimens. Here, we chose to generate a single average template since the biological variation was already known to be small; otherwise, specific templates would have been used to accommodate for larger differences. For example, such a registration-based method shows limitations to properly align specimens with a visible size difference compared to the average template. Specimens with observable differences in size should be categorized together, and specific templates should be built for each size category (e.g. Zamyadi et al. 2010; Wong et al. 2015). However, this was not necessary for the samples studied here.

Third, the resulting average atlas was manually segmented using 3DSlicer (Fedorov et al. 2012) http://www.slicer.org, and a labelmap segmenting the skull volume separately from both lower jaws was produced. Since the ANTs and ANTsR image

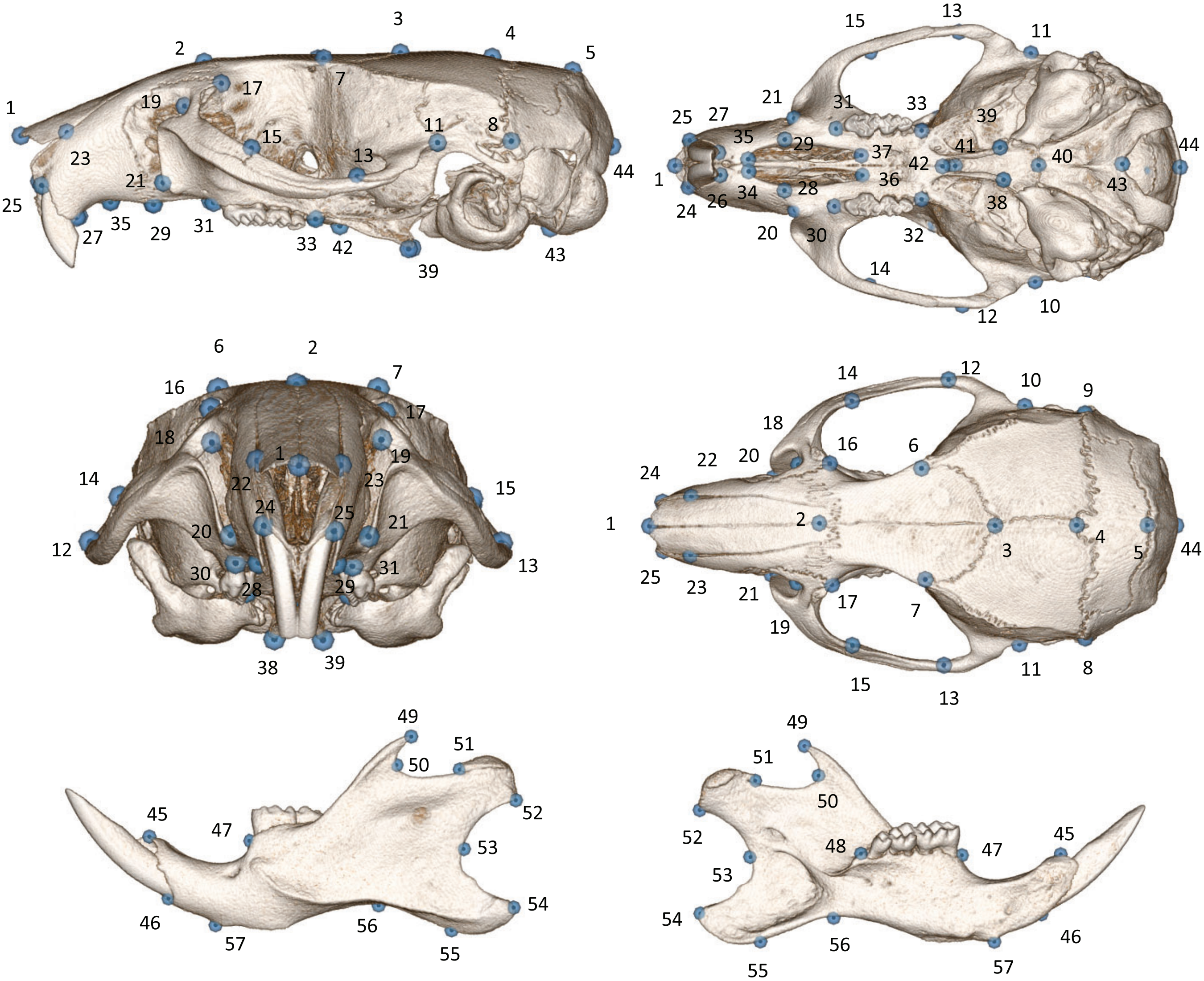

Fig. 1. Landmarks used for 3D phenotyping. For the detailed description of the landmarks, see Supplementary Table 1.

processing both rely on symmetric image deformation and normalization, it was deemed suitable to treat both lower jaws as a single unit that has bilateral symmetry, similarly as for the skull. Tests considering one lower jaw at a time produced unreliable automatic landmarking on such half structures (not shown here). The average template was landmarked using the same 70 landmarks as Pallares et al. (2015; Fig. 1, Supplementary Table 1) and twice to further test for measurement error.

Fourth, the original approach using ANTsR from Maga et al. (2017) was adapted to extend automatic landmarking to the lower jaws as well. However, the specimens were not registered onto the average template via the function inVariant(), since this function includes an additional rotation step designed for data with highly misaligned specimens and using this function caused misregistration with our data known to contain specimens already close to each other. Instead, the simpler default function antsRegistration() was used here. Prior to segmentation, quick checks were made from 2D slices using function plot.antsImage() to assess each specimen's overall degree of registration relative to the average template. Once this was confirmed, automatic segmentation was carried out to separate the skull from both lower jaws. Thereafter, a suite of image deformation and normalization was used separately for each segmented structure and for all specimens relative to the average atlas. All image transformations generated in the previous step were then applied to propagate the landmark configuration digitized onto the average template onto each segmented structure and for all specimens via function antsApplyTransformsToPoints(). All automatically placed landmarks for each structure were compiled into a single dataset for further geometric morphometric analyses.

Geometric morphometric analyses

Since the skull is an instance of object symmetry and the lower jaws exhibit matching symmetry, the original skull landmark configuration was duplicated, then the paired landmarks from this copy were swapped (relabeling). All landmark configurations from the left lower jaw were flipped (by multiplying all x coordinates by -1) to correspond to the landmark configurations digitized on the right lower jaw (Mardia et al. 2000; Kent and Mardia 2001).

A generalized Procrustes analysis (GPA) applied to the full landmark dataset extracted shape by removing extraneous effects of location, orientation, and position (e.g. Dryden and Mardia 1998). Thereafter, the Procrustes coordinates were averaged by individual to extract the component of symmetric shape variation while discarding the asymmetry. A principal component analysis was used on the covariance matrix of the Procrustes coordinates to extract the PC scores later used as shape data in subsequent analyses. Centroid size, the most common and explicit measure of size in geometric morphometrics, was computed as the square root of the sum of the squared distances of all landmarks from their centroid (e.g. Slice et al. 1996).

We assessed the effects of age and areal Bone-mineral density (aBMD) since there is a known predisposition of CFW mice toward abnormally high aBMD on shape (Pallares et al. (2015)) by multivariate regressions using 10,000 rounds of permutations separately for the skull and lower jaws as done in Pallares et al. (2015). For the skull, there was no association between age and shape ($r^2 = 0.00053$, $P = 0.75$) and a subtle but significant association between aBMD and shape ($r^2 = 0.00561$, $P = 0.033$). For the lower jaws, there were subtle but significant associations between age and shape ($r^2 = 0.00628$, $P = 0.0285$) and between aBMD and shape ($r^2 = 0.00608$, $P = 0.0283$). Given these weak associations, aBMD and age were not used as covariates in the subsequent trait mapping.

All analyses were done with the functions "gpagen()" for the GPA, "bilat.symmetry()" for object symmetry, and "procD.lm()" for the multivariate regressions from the R package geomorph (Adams and Otárola-Castillo 2013).

The average atlas was landmarked twice to test for digitizing error via the traditional 2-way mixed model Procrustes ANOVA and ANOVA used in studies of fluctuating asymmetry (FA; Leamy 1984; Palmer and Strobeck 1986; Savriama and Klingenberg 2011; Savriama et al. 2017). FA refers to small random differences between left and right sides of bilaterally symmetric structures and is, therefore, the smallest level of biological variation to which digitizing error is compared. Analysis was done in MorphoJ (Klingenberg 2011).

Genome-wide analyses

For QTL mapping, we use the same approach as in Pallares et al. (2015). Briefly, PCs summarizing at least 1% of variance were used for a univariate QTL mapping, and centroid size was considered for size QTL mapping. Each PC was analyzed separately.

Genome scans were conducted via the linear mixed model approach in GEMMA version 0.98.1 (Zhou and Stephens 2012) while accounting for the relationships among individuals (kinship matrix) and using the "leave one chromosome out" (LOCO) method in which a kinship matrix is calculated using markers from all other chromosomes except the ones that are on the chromosome under consideration (Cheng et al. 2013; Parker et al. 2014). We used the P-value computed from the likelihood ratio test output from GEMMA as the association test statistic. Permutation tests (1,000 rounds) were used to define genome-wide significance thresholds separately for each of the phenotypes used in the mapping. The distribution of minimum P-values was obtained from 1,000 permutations of the individual phenotypes while genotypes remained unchanged and the 95th percentile of this distribution was used as the significance threshold (Pallares et al. 2015). An average significance threshold per structure was computed including only phenotypes for which a significant association was found. Inflation of false positives might occur in this framework given that it does not allow samples to be swapped (Abney et al. 2002; Abney 2015); however, its robustness has been tested in many experiments involving advanced intercross lines which have inherent complex properties of relatedness (Parker et al. 2014).

Data from Pallares et al. (2015) were also reanalyzed with the same pipelines for further detailed comparisons between the two landmarking methods. The analysis confirmed the previous findings.

In addition, the function scanoneShape() from the R package shapeQTL (Navarro 2015) was used for multivariate mapping (Maga et al. 2015; Navarro and Maga 2016) using the same PCs included in the univariate mapping to further compare both landmarking methods. Shape is inherently multivariate, hence the use of this approach which simultaneously includes all PCs. Permutation tests (1,000 rounds) were used to define single genome-wide significance thresholds separately for each structure. The distribution of minimum P-values was obtained from the genome scan run at each permutation of the individual phenotypes while genotypes remained unchanged. The 95th percentile of this distribution was used as the significance threshold (Maga et al. 2015).

Here, it was not possible to define QTL regions based on the linkage disequilibrium (LD) pattern around the significant SNPs due to a sparse LD signal for all peak SNPs identified via the automatic landmarking approach (not shown here). Since this sometimes was also the case in Pallares et al. (2015), Bayes credible intervals of QTL (Dupuis and Siegmund 1999; Sen and Churchill 2001; Manichaikul et al. 2006; Maga et al. 2015; Navarro and Maga 2016) were instead calculated for the peak SNPs found in both landmarking methods via function bayesint() in the R/qtl package (Broman et al. 2003).

Overlap of relevant QTL regions determined in both landmarking methods was assessed via functions IRanges() and findOverlaps() from the GenomicRanges R package (Lawrence et al. 2013).

The position of peak SNPs identified in the automatic method was visualized via function chromPlot() from the chromPlot R package (Oróstica and Verdugo 2016), and overlapping QTL regions were highlighted via the zoom-in method implemented in the function chromoMap() from the chromoMap R package (Anand 2019).

Results

Digitizing error

The ANOVAs for centroid size and shape for skull and lower jaws both indicate that the "Individual-by-Side" (FA) interaction is highly significant ($P < 0.001$), which means that the smallest biological variation that can be measured here greatly exceeds the measurement error due to "Digitizing" (Supplementary Table 2). Note that there is no centroid size asymmetry for the skull since a unique configuration was considered for the whole structure and due to geometric constraints imposed by the GPA.

QTL mapping and overlap of QTL regions

For the atlas-based method, we kept all PCs summarizing at least 1% of the total variance as in Pallares et al. (2015), which together account for more than 99% of the variance of the skull and lower jaws shape. In Pallares et al. (2015), 22 PCs accounting for 84% of skull shape variation and 21 PCs representing 94% of mandible shape variation were used in the univariate mapping method, and the same PCs were used in the multivariate mapping as well. The average genome-wide significance thresholds were calculated separately for each analysis (see *Materials and Methods*) and are provided in the respective figure legends.

In the univariate mapping of the skull, none of the significant SNPs found in Pallares et al. (2015) were recovered using the automatic landmarking approach. However, 2 new significant SNPs were found with the automatic method (Fig. 2b). One overlaps with a QTL around the EGF pathway gene *Gab1* on chromosome 8, which was also found in Pallares et al. (2015) but for a different SNP marker, as depicted in Fig. 2g.

For the univariate mapping of the lower jaws, none of the 9 significant SNPs in 8 QTLs found in Pallares et al. (2015) were recovered using the automatic landmarking approach (Fig. 2, c and d). However, an overlap with 2 new significant SNPs was identified for a QTL around the transcriptional activator *Mn1* on chromosome 5 for the markers “rs33523792” and “rs47379127,” which have their QTL region overlapping within the one defined for the corresponding markers “rs33217671” and “rs33614268” found in Pallares et al. (2015; Fig. 2f).

For centroid size, no QTL was found for the skull, and 1 QTL on chromosome 1 was identified for the lower jaws. This is consistent with the results reported in Pallares et al. (2015). For the lower jaws, the same peak SNP was recovered by the automatic method (Fig. 2e).

In the multivariate mapping, none of the 4 skull QTLs found with the semiautomatic landmark data were recovered using the automatic landmarking approach, and the only marker “cfw – 5 – 46967657” identified on chromosome 5 does not have its QTL region overlapping with the other marker “rs32067860” defined on the same chromosome with the semiautomatic landmark data (Fig. 3, a and b). For the lower jaws, 2 QTLs were discovered using the automatic landmarking approach with the same major SNP “rs29385180” that was also found using the semiautomatic landmark data, which was the sole QTL found in this case (Fig. 3, c–e).

Discussion

Our results suggest that the accuracy of the automatic method for detecting variation required for QTL mapping is compromised, regardless of the mapping method used, though somewhat depending on the structure analyzed. For the skull that reflects a structure of a 3D organization, we found that most of the informative variance was lost, given the failure to detect the majority of QTLs from the previous study of Pallares et al. (2015). The informative variance of the lower jaw, which represents an object that is more 2D-like, where its outline can almost fully characterize the shape, was more adequately captured by the automatic method in either mapping technique, but still with less resolution than for the semiautomatic method. For centroid size of the lower jaws, the automatic method recovers the same QTL and the same SNP as the semiautomatic method. These identical results in both methods for this particular trait could be related to the fact that such a univariate measure of size seems easier to capture than a multivariate trait such as shape.

Our results do not appear to concur with the current trend accepting that such automatic method seems to be a reliable and promising tool, despite the known and quite often systematic differences between manual and automatic landmarking often evaluated either by linear distances and analyses of differences between group means and covariance matrices (Maga et al. 2017; Percival et al. 2019; Porto et al. 2021). In these particular studies, despite the known fact that the automatic landmarking procedure reduces the variance given its framework, the biological

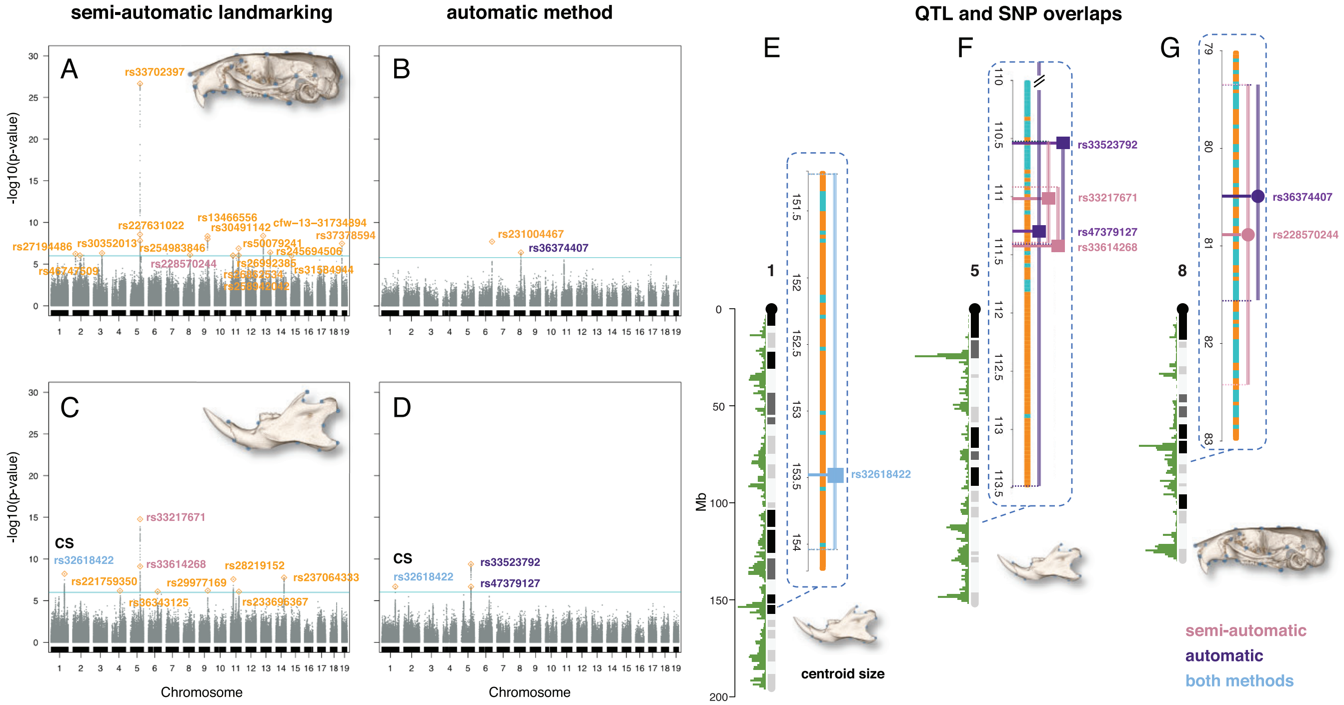

Fig. 2. Univariate mapping and locations of SNPs. Genome-wide scans for semiautomatic landmarking with manual adjustment (reanalysis of data from Pallares et al. 2015) and automatic landmarking (this study). a, b) For the skull and c, d) for the lower jaws. e–g) Overlapping QTL regions highlighted via zoom-in with 95% Bayesian credible intervals are indicated by lines from either side of each QTL. Cyan: genome-wide significance thresholds: −log10(P-value) = 6 for both the skull and lower jaws using the semiautomatic landmark data and 5.78 for the skull, and 6.04 for the lower jaws using the automatic landmark data. Orange: nonoverlapping QTL regions. Fuchsia and Purple: markers with overlapping QTL regions. Blue: the same marker found in both methods. Square: markers for the lower jaws. Circle: markers for the skull. CS, centroid size. Marker positions and statistics are provided in Supplementary Table 3.

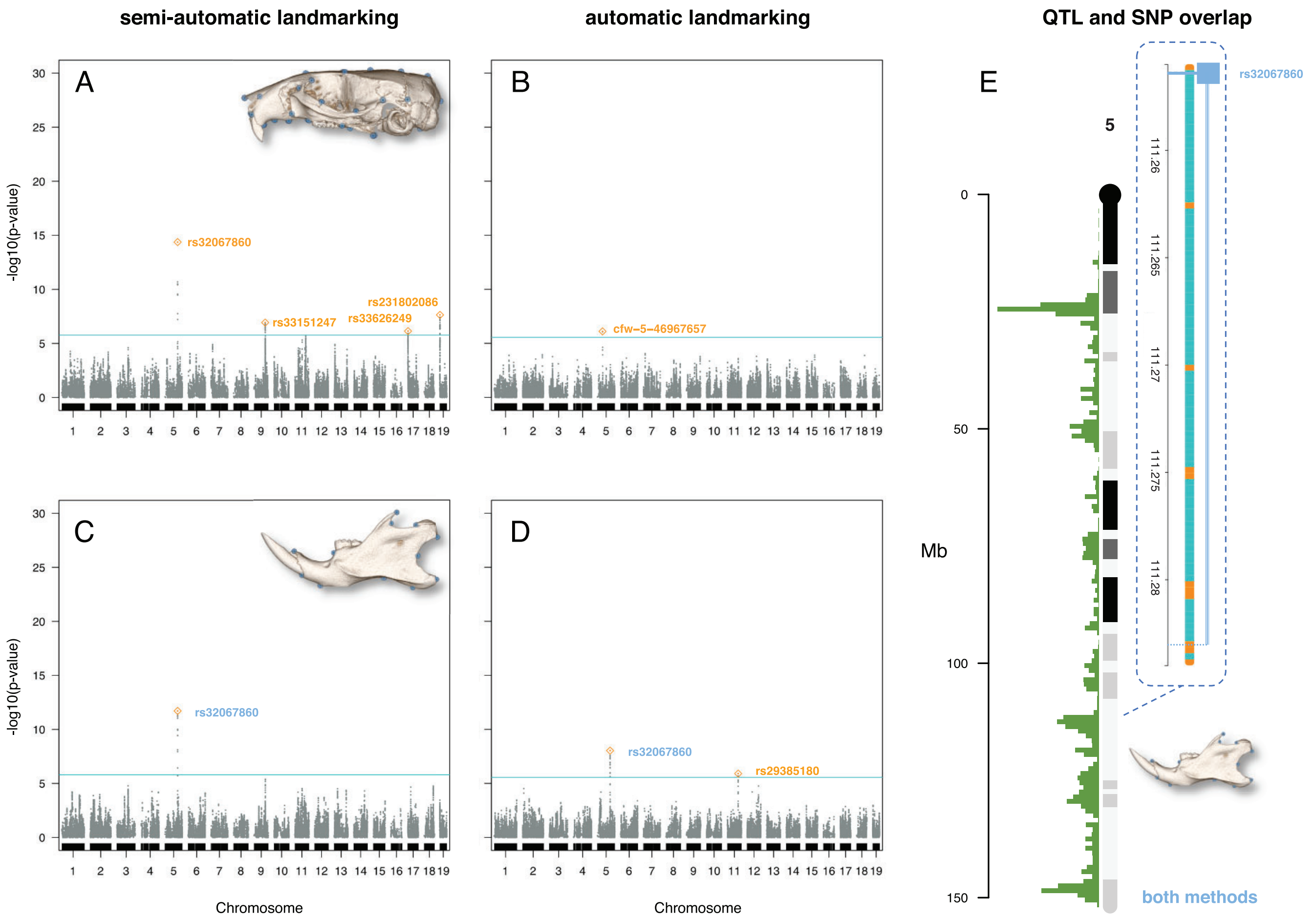

Fig. 3. Multivariate mapping and locations of SNPs. Genome-wide scans for semiautomatic landmarking with manual adjustment (reanalysis of data from Pallares et al. 2015) and automatic landmarking (this study). a, b) For the skull and c, d) for the lower jaws. e) Overlapping QTL regions highlighted via zoom-in with 95% Bayesian credible intervals are indicated by lines from either side of each QTL. Cyan: genome-wide significance thresholds: −log10(P-value) = 5.77 for the skull and 5.80 for the lower jaws using the semiautomatic landmark data and 5.56 for the skull, and 5.50 for the lower jaws using the automatic landmark data. Orange: nonoverlapping QTL regions. Blue: the same marker found in both methods. Square: markers for the lower jaws. Circle: markers for the skull. Marker positions and statistics are provided in Supplementary Table S3.

signal of interest is still captured via this method since shape differences between inbred mouse genotypes are known to be rather large. The atlas-based automatic method that relies on a suite of geometric transformations and image deformations to backpropagate the template's landmark configuration onto each specimen captures the actual biological signal only to some extent. Given that there is not much difference between mapping techniques, the automatic method causes this departure from the results obtained with the manual approach.

Here, we have illustrated and tested how the known differences between the landmarking approaches affect analyses. Our results suggest strong cautions regarding using such atlas-based automatic landmarking in geometric morphometric analyses, especially in genomewide association studies (GWAS), where the biological signal of interest is relatively small and cannot be adequately captured by it. Such mapping analyses can, in turn, be considered as an accurate and convenient tool for testing landmarking precision whenever applicable.

Ethics statement

No new animal experiments have been carried out since we reused CT scans that have been already produced in the reference study (Pallares et al. 2015).

Data availability

The complete code reproducing the analyses in Pallares et al. (2015) as well as in this study and data from Pallares et al. (2015) have been deposited at http://dx.doi.org/10.5061/dryad.k543p (last accessed Dec. 2021). Data generated for this study are available in the Supplementary data.

The software used in this study for template building and the R code for automatic segmentation and landmarking are freely available at:

- https://github.com/ANTsX/ANTs
- https://github.com/ANTsX/ANTsR
- https://github.com/muratmaga/mouse_CT_atlas

Supplemental material available at G3 online.

Acknowledgments

We thank Abraham Palmer for allowing the reanalysis of CT scans of mice generated by the Max-Planck Institute for Evolutionary Biology (Plön, Germany) that were part of a project supported by his National Institute of Health grant (R01GM097737) received from the University of California San Diego. We thank Nicolas Navarro and Murat Maga for their constructive remarks and comments on our work and Sylvain Gerber for stimulating discussions. We also thank Derk Wachsmuth, Kristian Ullrich, and Werner Wegner from our IT services and "Wallace" High Performance Computing Cluster, Carsten Fortmann-Grote from the Scientific Computing Unit at the MPI for Evolutionary Biology, the Max-Planck computing facilities at the Gesellschaft für Wissenschaftliche Datenverarbeitung Göttingen (GWDG), and the Max-Planck Computing and Data Facility (MPCDF).

YS and DT conceived the project and designed the analyses. YS processed the CT scans, collected and analyzed the data. YS and DT wrote the manuscript. Both authors read and approved the final manuscript.

Conflicts of interest statement

The authors declare no conflict of interest.

Literature cited

Abney M. Permutation testing in the presence of polygenic variation. Genet Epidemiol. 2015;39(4):249–258.

Abney M, Ober C, McPeek MS. Quantitative-trait homozygosity and association mapping and empirical genomewide significance in large, complex pedigrees: fasting serum-insulin level in the Hutterites. Am J Hum Genet. 2002;70(4):920–934.

Adams DC, Otárola-Castillo E. geomorph: an R package for the collection and analysis of geometric morphometric shape data. Methods Ecol Evol. 2013;4(4):393–399.

Anand L. 2019. chromoMap: an R package for interactive visualization and annotation of chromosomes. bioRxiv 605600. https://doi.org/10.1101/605600

Avants BB, Kandel BM, Duda JT, Cook PA, Tustison NJ. 2015. AntsR: an R package providing ANTs features in R. [accessed 2020 Oct 15]. https://github.com/ANTsX/ANTsR.

Avants BB, Tustison N, Song G. Advanced normalization tools (ANTS). Insight J. 2009;2(365):1–35.

Avants BB, Tustison NJ, Song G, Cook PA, Klein A, Gee JC. A reproducible evaluation of ANTs similarity metric performance in brain image registration. Neuroimage. 2011;54(3):2033–2044.

Broman KW, Wu H, Sen Ś, Churchill GA. R/qtl: QTL mapping in experimental crosses. Bioinformatics. 2003;19(7):889–890.

Bromiley PA, Schunke AC, Ragheb H, Thacker NA, Tautz D. Semi-automatic landmark point annotation for geometric morphometrics. Front Zool. 2014;11(1):1–21.

Burgio G, Baylac M, Heyer E, Montagutelli X. Genetic analysis of skull shape variation and morphological integration in the mouse using interspecific recombinant congenic strains between C57BL/6 and mice of the *Mus spretus* species. Evolution. 2009;63(10):2668–2686.

Cheng R, Parker CC, Abney M, Palmer AA. Practical considerations regarding the use of genotype and pedigree data to model relatedness in the context of genome-wide association studies. G3 (Bethesda). 2013;3(10):1861–1867.

Devine J, Aponte JD, Katz DC, Liu W, Lo Vercio LD, Forkert ND, Marcucio R, Percival CJ, Hallgrímsson B. A registration and deep learning approach to automated landmark detection for geometric morphometrics. Evol Biol. 2020;47(3):246–259.

Dryden IL, Mardia KV. 1998. Statistical Shape Analysis. Chichester: J. Wiley.

Dupuis J, Siegmund D. Statistical methods for mapping quantitative trait loci from a dense set of markers. Genetics. 1999;151(1):373–386.

Fedorov A, Beichel R, Kalpathy-Cramer J, Finet J, Fillion-Robin J-C, Pujol S, Bauer C, Jennings D, Fennessy F, Sonka M, *et al.* 3D Slicer as an image computing platform for the Quantitative Imaging Network. Magn Reson Imaging. 2012;30(9):1323–1341.

Frost SR, Marcus LF, Bookstein FL, Reddy DP, Delson E. Cranial allometry, phylogeography, and systematics of large-bodied papionins (primates: Cercopithecinae) inferred from geometric morphometric analysis of landmark data. Anat Rec A Discov Mol Cell Evol Biol. 2003;275(2):1048–1072.

Houle D, Govindaraju DR, Omholt S. Phenomics: the next challenge. Nat Rev Genet. 2010;11(12):855–866.

Katz DC, Aponte JD, Liu W, Green RM, Mayeux JM, Pollard KM, Pomp D, Munger SC, Murray SA, Roseman CC, *et al.* Facial shape and allometry quantitative trait locus intervals in the Diversity Outbred mouse are enriched for known skeletal and facial development genes. PLoS One. 2020;15(6):e0233377.

Kent JT, Mardia KV. Shape, Procrustes tangent projections and bilateral symmetry. Biometrika. 2001;88(2):469–485.

Klingenberg CP. MorphoJ: an integrated software package for geometric morphometrics. Mol Ecol Resour. 2011;11(2):353–357.

Lawrence M, Huber W, Pagès H, Aboyoun P, Carlson M, Gentleman R, Morgan MT, Carey VJ. Software for computing and annotating genomic ranges. PLoS Comput Biol. 2013;9(8):e1003118.

Leamy L. Morphometric studies in inbred and hybrid house mice. V. Directional and fluctuating asymmetry. Am Nat. 1984;123(5): 579–593.

Maga AM, Navarro N, Cunningham ML, Cox TC. Quantitative trait loci affecting the 3D skull shape and size in mouse and prioritization of candidate genes in-silico. Front Physiol. 2015;6:92.

Maga AM, Tustison NJ, Avants BB. A population level atlas of *Mus musculus* craniofacial skeleton and automated image-based shape analysis. J Anat. 2017;231(3):433–443.

Manichaikul A, Dupuis J, Sen S, Broman KW. Poor performance of bootstrap confidence intervals for the location of a quantitative trait locus. Genetics. 2006;174(1):481–489.

Mardia KV, Bookstein FL, Moreton IJ. Statistical assessment of bilateral symmetry of shapes. Biometrika. 2000;87(2):285–300.

Meijering EH, Niessen WJ, Viergever MA. Quantitative evaluation of convolution-based methods for medical image interpolation. Med Image Anal. 2001;5(2):111–126.

Muñoz MM, Price SA. The future is bright for evolutionary morphology and biomechanics in the era of big data. Integr Comp Biol. 2019;59(3):599–603.

Navarro N. 2015. shapeQTL: shape QTL mapping experiment with R. [accessed 2020 Oct 15]. https://github.com/nnavarro/shapeQTL.

Navarro N, Maga AM. Does 3D phenotyping yield substantial insights in the genetics of the mouse mandible shape? G3 (Bethesda). 2016;6(5):1153–1163.

Oróstica KY, Verdugo RA. chromPlot: visualization of genomic data in chromosomal context. Bioinformatics. 2016;32(15):2366–2368.

Pallares LF, Carbonetto P, Gopalakrishnan S, Parker CC, Ackert-Bicknell CL, Palmer AA, Tautz D. Mapping of craniofacial traits in outbred mice identifies major developmental genes involved in shape determination. PLoS Genet. 2015;11(11):e1005607.

Pallares LF, Turner LM, Tautz D. Craniofacial shape transition across the house mouse hybrid zone: implications for the genetic architecture and evolution of between-species differences. Dev Genes Evol. 2016;226(3):173–186.

Palmer AR, Strobeck C. Fluctuating asymmetry: measurement, analysis, patterns. Annu Rev Ecol Syst. 1986;17(1):391–421.

Parker CC, Carbonetto P, Sokoloff G, Park YJ, Abney M, Palmer AA. High-resolution genetic mapping of complex traits from a combined analysis of F2 and advanced intercross mice. Genetics. 2014;198(1):103–116.

Pavlicev M, Mitteroecker P, Gonzalez PN, Rolian C, Jamniczky H, *et al.* Development shapes a consistent inbreeding effect in mouse crania of different line crosses. J Exp Zool B Mol Dev Evol. 2016;326:474–488.

Percival CJ, Devine J, Darwin BC, Liu W, van Eede M, Henkelman RM, Hallgrimsson B. The effect of automated landmark identification on morphometric analyses. J Anat. 2019;234(6):917–935.

Porto A, Rolfe SM, Maga AM. ALPACA: a fast and accurate approach for automated landmarking of three-dimensional biological structures. Methods Ecol Evol. 2021;12:2129–2144.

Savriama Y, Gerber S, Baiocco M, Debat V, Fusco G. Development and evolution of segmentation assessed by geometric morphometrics: the centipede *Strigamia maritima* as a case study. Arthropod Struct Dev. 2017;46(3):419–428.

Savriama Y, Klingenberg CP. Beyond bilateral symmetry: geometric morphometric methods for any type of symmetry. BMC Evol Biol. 2011;11(1):280.

Schindelin J, Arganda-Carreras I, Frise E, Kaynig V, Longair M, Pietzsch T, Preibisch S, Rueden C, Saalfeld S, Schmid B, *et al.* Fiji: an open-source platform for biological-image analysis. Nat Methods. 2012;9(7):676–682.

Sen S, Churchill GA. A statistical framework for quantitative trait mapping. Genetics. 2001;159(1):371–387.

Slice D, Bookstein F, Marcus L, Rohlf F. Appendix I: a glossary for geometric morphometrics. Nato ASI Series A Life Sci. 1996;284:531–552.

Tustison NJ, Cook PA, Holbrook AJ, Johnson HJ, Muschelli J, Devenyi GA, Duda JT, Das SR, Cullen NC, Gillen DL, *et al.* The ANTsX ecosystem for quantitative biological and medical imaging. Sci Rep. 2021;11(1):1–13.

Tustison NJ, Shrinidhi KL, Wintermark M, Durst CR, Kandel BM, Gee JC, Grossman MC, Avants BB. Optimal symmetric multimodal templates and concatenated random forests for supervised brain tumor segmentation (simplified) with ANTsR. Neuroinformatics. 2015;13(2):209–225.

Weiss K, Buchanan A, Richtsmeier J. How are we made?: Even well-controlled experiments show the complexity of our traits. Evol Anthropol. 2015;24(4):130–136.

Williams G. 2005. NIfTI input/output. [accessed 2020 Oct 15]. https://imagej.nih.gov/ij/plugins/nifti.html.

Wong MD, van Eede MC, Spring S, Jevtic S, Boughner JC, Lerch JP, Henkelman RM. 4D atlas of the mouse embryo for precise morphological staging. Development. 2015;142(20):3583–3591.

Zamyadi M, Baghdadi L, Lerch JP, Bhattacharya S, Schneider JE, Henkelman RM, Sled JG. Mouse embryonic phenotyping by morphometric analysis of MR images. Physiol Genomics. 2010;42A(2): 89–95.

Zhou X, Stephens M. Genome-wide efficient mixed-model analysis for association studies. Nat Genet. 2012;44(7):821–824.

Communicating editor: F. P.-M. de Villena