Variation und Variabilität der Unterkieferform in der Hausmaus

Dissertation

zur Erlangung des Doktorgrades
der Mathematisch-Naturwissenschaftlichen Fakultät
der Christian-Albrechts-Universität zu Kiel
vorgelegt von

Louis A. Boell

Plön, 2010

Referent: Prof. Diethard Tautz

Koreferent: Prof. Thomas Bosch

Tag der mündlichen Prüfung: 13.8.2010

Zum Druck genehmigt: Kiel, 13.8.2010

gez. der Dekan

Hiermit erkläre ich, daß die vorliegende Arbeit

- nach Inhalt und Form meine eigene ist, abgesehen von der Beratung durch meinen Betreuer Prof. Diethard Tautz;
- daß diese Arbeit an keiner anderen Stelle zur Promotion vorgelegen hat oder vorliegt und auch noch nicht veröffentlicht ist;
- daß diese Arbeit unter Einhaltung der Regeln guter wissenschaftlicher Praxis der Deutschen Forschungsgemeinschaft entstanden ist.

Plön, den 7.6.2010

Louis Boell

Danksagung

Nicht nur im fachlichen Sinn möchte ich mich besonders bei meinem Betreuer, Prof. Diethard Tautz, bedanken, der meiner Eigenwilligkeit viel Raum gelassen hat, mich gleichzeitig aber materiell und durch umfassende Diskussionen großzügig unterstützt hat und dessen konstruktive Kritik und intellektuelle Beiträge meine Arbeit und mein Verständnis entscheidend befördert haben. Außerdem gilt mein Dank vielen Kollegen, deren Teilnahme an meinen den ihren oft fernerstehenden Forschungsinhalten meine Arbeit bereichert hat. Heike Harre, Christine Pfeifle und Maren Volquardsen haben mir bei der Arbeit mit den Mäusen viel geholfen. Anja Schunke möchte ich für eine wunderbare, jahrelange Bürogemeinschaft danken.

Da die Dissertation den Abschluß einer langwierigen Ausbildungs- und Entwicklungsphase darstellt, möchte ich diese Gelegenheit benutzen, um meinen Respekt und meine Dankbarkeit gegenüber einigen mir nahestehenden Menschen auszudrücken, deren Einfluß für mich maßgeblich war und ist.
An erster Stelle möchte ich meine Mutter Anemone Boell nennen, an zweiter Stelle Birgit Hower.
Des weiteren möchte ich danken: Mathilde Boell, Wolfgang Blaicher, Ulrich Burkhart, Christine Elbel, Alex Franzke, Moritz Hasselgruber, Jan Lukas, Farid Mirghawameddin, Heiko Muth, Sven Petersen, Hendrik Schmidt, Rolf Stabel, Niklas Wacholt, Simone Weyers.

Schließlich, stellvertretend für einen großen Reichtum an kostbaren Einflüssen und Gaben: Drexciya.

Table of contents

Wer will was Lebendigs erkennen und beschreiben,
Sucht erst den Geist heraus zu treiben,
Dann hat er die Teile in seiner Hand,
Fehlt, leider! nur das geistige Band.

J. W. v. Goethe, Faust I

Zusammenfassung

In dieser Doktorarbeit beschreibe ich zunächst die Variation in der Unterkieferform wildgefangener Mäuse, wobei das Hauptaugenmerk auf der Hausmaus, *Mus musculus*, liegt. Unter Einbeziehung gefangengehaltener Mäuse, von Inzuchtstämmen und einiger Experimentalpopulationen versuche ich dann herauszuarbeiten, welche biologischen Prozesse die beobachteten Variationsmuster erklären könnten. Hierbei kommen auch genetische und entwicklungsbiologische Aspekte zum Tragen, d.h. die der Variation zugrundeliegende Variabilität. Meine wichtigsten Ergebnisse sind folgende:

1) Mahalanobis-Distanzen basierend auf canonical-variates-Analyse von Prokrustes-Koordinaten sind ein gutes Maß für den Formunterschied zwischen zwei Populationen. In Kombination mit der Benutzung mikrotomographischer Aufnahmen von Mäuse-Hemimandibeln sind sie ausreichend robust gegenüber verschiedenen Problemen betreffend Qualität der Proben und Weiterverarbeitung der Daten, als da sind Begrenzungen der Probengröße, systematische Fehler in Bezug auf Alter, Geschlecht und Größe der Tiere, Orientierung der Specimen im Tomographen, Präparation der Knochen und Registrierung der Meßpunkte.
2) Phänotypische Plastizität als Reaktion auf Umweltfaktoren beeinflußt die Unterkieferform weniger stark, als eß dem durschschnittlichen Formunterschied zwischen zwei Populationen entspricht, d.h. Formunterschiede zwischen wilden Populationen beruhen zu großen Teilen auf genetischen Unterschieden.
3) Verschiedene Kategorien von Selektion könnten auf die Unterkieferform gewirkt haben. Vier Populationen von Mäusen aus sommertrockenen Gebieten haben sehr ähnliche Formen, was bedeuten könnte, daß diese Form durch stabilisierende Selektion konserviert wurde. Eine Population von *M. m. domesticus* aus einem Dorf in den spanischen Pyreneen, wo die Mäuse sympatrisch mit *M.spretus* lebten, ist in ihrer Unterkieferform stark divergent von andern *M. musculus*. Es könnte sich hierbei um einen Fall von character displacement, d.h. evolutionäre Betonung der Nischenunterschiede zur Konkurenzvermeidung, handeln. Zwei Populationen von *M. m. domesticus*, die sich von nicht weit zurückliegenden Kolonisationsereignissen durch menschlichen Transport auf den subantarktischen Kerguelen-Inseln herleiten, weichen von anderen Hausmäusen in teilweise ähnlichen Richtungen ab, was einen Fall paralleler Anpassung an das kalte Klima auf diesen Inseln darstellen könnte.

4) Inzuchtstämme der Hausmaus unterscheiden sich stärker von wilden Hausmäusen und voneinander als unterschiedliche Arten in der Natur. Dies beruht vermutlich auf nichtadditiven, epistatischen Interaktionen zwichen and der Morphogenese beteiligten Genen. Diese Hypothese wird unterstützt durch die Beobachtung, daß F1-Tiere aus Kreuzungen zwischen verschiedenen Inzuchtstämmen nicht lediglich wie Zwischeformen zwischen den Parentalstämmen aussehen, sondern sich wieder teilweise der Wildform annähern.
5) Genetisch diverse (nicht ingezüchtete) Populationen von wilden Hausmäusen, die im Labor gehalten werden, verändern ihre Unterkieferform im Lauf weniger Generationen, wenn auch weniger stark als Inzuchtstämme. Sie unterscheiden sich auch weniger voneinander als dies bei Inzuchtstämmen der Fall ist. Möglicherweise liegt hier ein noch unbekannter (epigenetischer) Mechanismus zugrunde, der durch die Laborhaltung induziert wird.
6) Die Formabweichungen von Kerguelen-Mäusen, Inzuchtstämmen und gefangen gehaltenen Wildmäusen („abgeleitete Populationen“) sind in ihren Richtungen nicht zufallsverteilt. Dieses Muster muß aus einer merkmalsbezogenen Perspektive untersucht werden. Die geometrische Morphometrie bietet dafür keine geeigneten Methoden. Einfache Alternativmethoden, basierend auf Meßstrecken, ermöglichen es, die Ähnlichkeit der Richtungen von Formveränderungen zu quantifizieren und die beteiligten Regionen des Unterkiefers zu identifizieren.
7) Mithilfe eines eigens entwickelten manuellen Protokolls wurden 20 Gruppen miteinander kovariierender Meßstrecken (interlandmark distances, ILMDs) identifiziert, die wiederum in 5 „Hauptmerkmale“ gruppiert werden können. Diese resultieren möglicherweise aus unterschiedlicher Zuweisung von begrenzten Wachstumsressourcen zu den Fortsätzen des Unterkiefers oder aus funktioneller Koppelung zwischen voderen und hinteren Bereichen.
8) Die 5 Hauptmerkmale „erklären“ große Anteile der Variation in verschiedenen Zusammenhängen: Abweichung der „abgeleiteten Populationen“ von Wildmäusen, Variation innerhalb sowohl genetisch diverser als auch ingezüchteter Populationen (im letzteren Fall besteht ein Zusammenhang zur Instabilität von Entwicklungsvorgängen), „epistatische Abweichungen“ bei Auszuchttieren vom Mittelwert zwichen den Elternstämmen, sowie nachgeburtliche Formveränderungen. Diese Vielfalt von Zusammenhängen ist ein Hinweis darauf, daß ein Großteil genetischer und entwicklungsbiologischer Veränderungen sich in einer begrenzten Anzahl von

Formveränderungen niederschlägt. Die 5 Hauptmerkmale sind allerdings von geringerer Bedeutung für die Erklärung von Formunterschieden zwischen Populationen und Arten.

9) Unter Zusammenfassung der Hinweise auf die epistatische Grundlage bestimmter Formunterschiede und der spezifischen Zusammenhänge, in denen sich diese Formunterschiede manifestieren, schlage ich folgende Hypothese vor: epistatische Varianz und Instabilität der Entwicklung produzieren einen Großteil der Variation innerhalb von Populationen. Die Unterschiede zwichen Populationen und Arten, die durch genetischer Evolution entstehen, beruhen hauptsächlich auf additiver genetischer Varianz, die andere Formunterschiede hervorbringt als Epistasias und Instabilität. Diese Varianz wird möglicherweise durch stabilisierende Selektion eingeschränkt und spielt daher innerhalb von Populationen eine geringere Rolle.

Summary

In this thesis, I provide a description of the shape space of wild mouse mandibles with a focus on *Mus musculus*. Extending the comparisons to captive mice, inbred strains and some experimental populations, I try to infer which biological processes might account for observed patterns of shape variation, including genetic and developmental aspects (variability). I obtain the following results:

1) Mahalanobis distances based on CVA of Procrustes coordinates are a good measure of the global shape difference between two populations. Combined with the use of two-dimensional projections of µCT images of mouse hemimandibles, they are sufficiently robust in the face of diverse problems with sample quality and data processing, such as limitations in sample size, sampling errors with respect to sex, age and size of the animals, orientation of the samples inside the µCT scanner, preparation of bones and landmark digitization error.
2) Phenotypic plasticity as a reaction to environmental differences affects mandible shape by a smaller amount than the average distance between samples of wild-caught populations, suggesting that the shape differences between wild populations mostly have a genetic basis.
3) Various types of selection may have acted on shape. Four populations of mice from summer-dry regions cluster closely together, indicating that stabilizing selection may have conserved their shape. A *M. m. domesticus* sample from a site in Spain where the mice live in sympatry with a population of *M. spretus* is highly divergent from other *M. musculus*. This could represent a case of character displacement. Two populations of *M. m. domesticus* representing rather recent events of colonization on the subantarctic Kerguelen islands have diverged from other *M. musculus* in partially similar directions, which could represent an adaptation to the cold climate on these islands.
4) Inbred mouse strains are more divergent from wild mice and from each other than different species in nature, suggesting that nonadditive mechanisms of inheritance, especially epistasis, are important determinants of shape. This idea is supported by the finding that F1 of outcrosses between inbred strains look more similar to wild mice

than their parentals, i. e. their phenotype is not just intermediate, and there is some complementation of changes from the wildtype, but no complete reversal.

5) Wild-derived outbred populations kept in the laboratory diverge from wild mice over the course of a few generations, albeit less so than inbred mice. They are, however, not more divergent from each other than wild populations. This finding may point toward the existence of some epigenetically inherited mechanism of shape change which is somehow induced under laboratory conditions.
6) The Kerguelen mice, inbred strains, and wild-derived outbred populations ("derided populations") do not diverge from wild mice in random directions. This pattern needs to be analyzed from a trait-based perspective. Geometric morphometrics alone is not suitable to dissect overall variation into individual traits. Simple alternative methods based on interlandmark distances (ILMDs) help to quantify the similarity between directions of shape change and to dissect shape changes with respect to the mandibular subregions involved.
7) Using a purpose-designed manual protocol, 20 groups of covarying ILMDs are identified, which can themselves be grouped into 5 "major traits". These can largely be assumed to represent tradeoffs of tissue mass allocation during growth and to some degree functional coupling between parts of the mandible.
8) The 5 major traits explain large proportions of variation in several contexts: divergence of the derived populations from wild mice, variation within outbred and inbred populations (for the latter, i.e. developmental instability), "epistatic deviations" of outcross F1 from the interparental mean, and postnatal longitudinal ontogenetic shape change. This variety of contexts indicates that a large part of the genetically/developmentally generated variation is expressed via a limited number of types of shape changes. At the same time, the 5 traits are less important for the explanation of differences between populations and species.
9) Taking together the evidence for epistatic genetic architecture of shape and the results on the specific contexts in which the corresponding shape changes are observed, I hypothesize that epistatic shape variance may relate to developmental instability and provides the major part of phenotypic variance in wild populations. Stabilizing selection is unable to control this variation. Evolutionary divergence, however, happens predominantly along axes of additive variance, which provide a lower part of phenotypic variance within populations under this model, potentially due to the action of stabilizing selection.

Introduction

Origination and maintenance of phenotypic variation in natural populations constitute a largely unresolved problem in evolutionary biology. This is especially true for multivariate characters such as the shape of biological structures. The study of multivariate phenotypic evolution is different from the study of univariate traits in that correlations between traits have to be considered, which is more realistic for natural evolutionary processes than the univariate scenario (Lande 1979, Lande and Arnold 1983). This reflects the fact that a genome of limited information content has to orchestrate the evolution of an infinite number of measurable traits (Johnson and Barton 2005). The study of the evolution of shape is of special interest as a paradigm of multivariate phenotypic evolution, because morphometric traits within the same structure can be supposed to be closely related through genetics and development (Klingenberg 2009).

There have been many case studies of shape evolution in many different taxa, yet these either had their focus on macroevolutionary trends (e.g., Stayton 2005, Wroe 2007), or they were set in the context of comparisons between a few specific populations (e.g. Auffray and Latieule1996, Corti and Rohlf 2001, Renaud and Auffray 2009).

While these studies throw light on diverse aspects of shape evolution, they leave one fundamental issue unaddressed: the broader properties of shape variation within and between populations of the same species, i. e. the shape space of the species. This shape space, however, is fundamental for deriving the relevant evolutionary questions, very much in the same way as knowledge of comparative anatomy is relevant for the study of macroevolution. Thus, we need to know how large variation is within populations compared to between populations and between species, how phenotypic variation is distributed with respect to evolutionary history and environmental conditions, and what are the major directions of phenotypic variation. Only this knowledge will enable us to identify relevant study populations and to gauge the relevance and interpretation of more specific and experimental results. This point applies to a number of questions about the microevolutionary dynamics of shape:

- Is phenotypic plasticity sufficient for organisms to adapt their phenotypes to local environments? Does plasticity thus eliminate the need for further adaptive evolution? This question can only be answered by quantitatively comparing the effect of environmental effects on shape to the natural range of shape differences between populations of the same species.
- Are allopatric populations of the same species distinct at all? How much evolution does occur within species and how much of it could be due to random drift? How large is

variation within populations with respect to variation between populations and between species? How tightly might stabilizing selection control the phenotype?

- Which populations within a species are most distinct from other populations? Can any relationship with phylogeographic or environmental patterns be established?
- Are the major directions of variation similar between populations, indicating species-typical properties of the developmental system or of genetic architecture? Does the "lines of least resistance" concept (Schluter 1996), which predicts that the major directions of variation within populations should also be major directions of evolutionary divergence, apply within species?
- How different are captive and inbred animals from wild animals in general? Inbreeding implies a strong bottleneck. Bottlenecks cause strong drift and may change the structure of genetic variance (Turelli and Barton 2006). What might we learn from the differences associated with inbreeding about the genetic architecture of shape?

If these questions could be studied in a common dataset consisting of a sufficient number of population samples to represent the natural shape space of the species, it might become possible to draw for the first time a coherent picture of shape variation within a species and to identify the most important directions of further research on this topic. Linking these results with more specific experimental research should then become a powerful entry into the microevolution of shape.

The morphospace of mouse mandibles offers the perspective to address some aspects of this field of questions on a descriptive and potentially also on an experimental level. The mouse is on the one hand a cosmopolitan animal with a well-known evolutionary and phylogenetic history (Cucchi et al. 2005, Tucker et al. 2005), and on the other hand a highly tractable genetic model system, for which many relevant genetic resources have already been developed.

The mouse mandible is a complex structure whose shape is rich in information and its shape can be conveniently studied in two dimensions (2D), offering the possibility of rapidly phenotyping large numbers of animals as has been shown, e.g., in several QTL studies (e.g., Cheverud et al. 1997, 2004, Ehrich et al. 2003, Klingenberg and Leamy 2001, Klingenberg et al. 2001, 2003, 2004, Leamy et al. 1997, 2008,). Furthermore, a body of knowledge exists at least about the earlier (prenatal) developmental stages of the mouse mandible (Atchley and Hall 1991), and condylar growth is a well-studied model of craniofacial bone growth (Mao and Nah 2004).

Although there has been a wealth of evolutionary studies about mouse mandible shape (Auffray and Latieule 1996, Corti and Rohlf 2001, Davis 1983, Gerasimov et al. 1991, Macholan 1996a, 1996b, 2008, Macholan et al. 2008, Pergams and Ashley 2001, Renaud and Auffray. 2009, Scriven and

Bauchau1992, Slabova and Frynta 2007), the approach that has been taken in the present thesis provides for the first time a general description of the natural morphospace of house mouse mandibles based on museum material and additional wild-caught mouse samples, including also samples of two sister species. This makes it possible to draw general conclusions about the comparative intra- and interspecific shape variation, and to evaluate experimentally generated shape variation on the background of this general knowledge about natural variation.

Beyond the description of the natural morphospace, this thesis does take up several threads of discussion already present in the literature about evolution, development and genetics of mouse mandible shape:

- Phenotypic plasticity in mouse mandibles has been traditionally investigated using hard diet – soft diet experiments (e.g., Levrini et al. 2003, Maki et al. 2002, Renaud and Auffray 2009, Renaud et al. 2010, Yamada and Kimmel 1991). Here, this is complemented with a wild caught – laboratory comparison. The morphospace data make it possible for the first time to put the results into a quantitative comparative context.
- QTL studies have found substantial epistasis in the genetic architecture of skull and mandible shape (Burgio et al. 2009, Cheverud et al. 2004). In this thesis, epistasis is addressed as explanatory factor in the context of morphological changes associated with inbreeding.
- The mouse mandible has been hypothesized to evolve along “lines of least resistance” (Schluter 1996), i.e. in directions corresponding to the major axes of within-population variance in the context of “insular evolution” (Renaud and Auffray 2009). I evaluate this proposition on the broader comparative background of the natural morphospace.
- The dichotomy of genetic variation vs. developmental instability has been traditionally studied using fluctuating asymmetry (Leamy et al. 2005). By comparing amount and structure of the shape variance between outbred and inbred strains under laboratory conditions, we get an estimate of the amount and structure of shape variance occurring independently of genetic and environmental factors.

My thesis is thus an attempt to advance our understanding of the origination and maintenance of shape variation in the mouse mandible in the descriptive context of a broadly sampled shape space. The hope is that this might provide new perspectives on various problems associated with phenotypic variation and evolution.

Material and Methods

Samples

The morphospace study comprises 34 population and strain sample sets from mice of the subgenus *Mus* including wild-caught, captive and inbred mice (detailed in Table 1). A minimum of 14 individuals were scored for each sample set, i.e. being equal to or exceeding the number of landmarks scored.

Name	Species	Geographical origin	N
wild population samples			
DOM EGYPT	*M. m.domesticus*	An Nawamis, Middle Egypt	29
DOM IRAN AHVAZ	*M. m.domesticus*	Ahvaz, South Iran	19
DOM IRAN TEHERAN	*M. m.domesticus*	Teheran region, Iran	16
DOM SICILY	*M. m.domesticus*	Sicily	29
DOM SPAIN PUDEMONT	*M. m.domesticus*	Puente de montanana, Spain	26
DOM KERG GOUILLOU	*M. m.domesticus*	Kerguelen islands Gouillou	27
DOM KERG COCHONS	*M. m.domesticus*	Kerguelen islands Cochons	27
DOM GER MUNICH	*M. m.domesticus*	Munich, Germany	14
DOM GER FRANKFURT	*M. m.domesticus*	Frankfurt, Germany	23
MUS HUNGARY	*M. m. musculus*	Hungary	25
CAS JOHNSTON ATOLL	*M. m. castaneus*	Johnston Atoll	14
MAC GREECE	*M. macedonicus*	Chios, Greece	16
MAC TURKEY	*M. macedonicus*	Southwest Turkey	17
SPR SPAIN PUDEMONT	*M. spretus*	Puente de montanana, Spain	29
SPR SPAIN MADRID	*M. spretus*	Madrid, Spain	28
"classical" laboratory strains			
BALB/cByJ	mixed *M. musculus* ancestry	undefined origin	30
FVB/NJ	mixed *M. musculus* ancestry	undefined origin	30
C57BL/10J	mixed *M. musculus* ancestry	undefined origin	30
C57BL/6J	mixed *M. musculus* ancestry	undefined origin	18
wild derived inbred strains			
DOM WD INBRED STLT	*"M. m. domesticus"*	Straas, Germany	20
DOM WD INBRED STRA	*"M. m. domesticus"*	Straas, Germany	20
DOM WD INBRED STRB	*"M. m. domesticus"*	Straas, Germany	20
CAS WD INBRED CAST/EIJ	*"M. m. castaneus"*	Thonburi, Thailand	15
Mus wd inbred PWD	*"M. m. musculus"*	Kunratice, Czech Republic	20
wil-derived outbred populations			
DOM LAB IRAN AHVAZ GEN1	*M. m.domesticus*	Ahvaz, Iran	24
DOM LAB IRAN AHVAZ GEN3	*M. m.domesticus*	Ahvaz, Iran	15
DOM LAB FRANCE GEN 3/4	*M. m.domesticus*	Massif Central, France	17
DOM LAB GER COLOGNE	*M. m.domesticus*	Cologne, Germany	15
MUS LAB KHAZ	*"M. m. musculus"*	Region Almaty	15
CAS WD (INBRED) TAIWAN	*"M. m. castaneus"*	Taiwan	14
SPRE WD (INBRED) SPAIN MADRID	*M. spretus*	Madrid, Spain	15
outcrosses			

C57BL/6J x PWD F1	outcross	undefined origin	14
C57BKL/6J x CAST/EIJ F1	outcross	undefined origin	15
FRANCE x IRAN F1	outcross	undefined origin	94

Table1. Overview of the samples included in the shape space study. The remaining samples (for assessment of technical problems etc.) are detailed in the text.

Most sample sets come from free-living mice which were caught in the field (Table M.1.). They were obtained either as loans from natural history museums or directly from colleagues. Museum samples were taxonomically reanalysed, based on the character list provided by (Macholan 1996a). Only unequivocally assignable material was used. For example, the samples from Puente de Montanana (Spain) turned out to be a mixture between *M. musculus* and *M. spretus*, which live sympatrically in this region. To assess interspecific divergence, we included two samples of each *Mus macedonicus* and *Mus spretus*. These species (together with *M. spicilegus*) are the closest sister species of *M. musculus* (Tucker et al. 2005).

Five sample sets of wild-derived inbred strains were studied. Three of them (Stlt, StrA, StrB, Pialek et al. 2008; hereafter referred to as DOM WD INBRED) are derived from *M. m. domesticus* and were obtained from the breeding facility in Studenec (Czech Republic). Their average ages at dissection were 21 weeks (ranging from 10 to 72 weeks), 23 weeks (ranging from 16 to 45 weeks) and 20 weeks (ranging from 9 to 33 weeks), respectively. Samples of PWD (*M. m. musculus*; JAX stock number 0046660; age at dissection 10 to 12 weeks, hereafter referred to as MUS WD INBRED) and Cast/EiJ (*M. m. castaneus*; JAX stock number 000928; age at dissection 9 to 12 weeks, hereafter referred to as CAS WD INBRED) were from the breeding facilities at the MPI for Genetics in Berlin.

Four sample sets represent classical inbred laboratory mouse strains. Three of them (BALB/cByJ, FVB/NJ, C57BL/10J, hereafter referred to as LAB INBRED) were obtained from the Jackson laboratory (Bar Harbor, USA). These strains were chosen for their different mandible measurements as indicated in the mouse phenome database of Jackson laboratory (http://phenome.jax.org/pub-cgi/phenome/mpdcgi?rtn=projects/details&sym=Everett1). These mice were killed at 10 weeks of age, shipped frozen and thawed for scanning. An additional sample of LAB INBRED C57BL/6J mice (age at dissection 11 to 13 weeks) was obtained from the breeding facilities at the MPI for Genetics in Berlin.

Seven samples represent wild-derived outbred populations, i.e. populations of captive mice established from wild-caught mice from specific locations and held in captivity for one or more generations (see table 1). The DOM LAB IRAN AHVAZ GEN1 sample consists of

mandibles prepared using the detergent method from mice of mixed ages. All other samples in this category consist of mice scanned alive at 8 to10 weeks age.

Additional sample sets were used to assess the impact of technical problems and to investigate ontogeny and plasticity, as well as the phenotypic effect of outcrossing inbred strains.

For the comparative developmental series, juvenile and adult specimens of LAB INBRED C57Bl/6J and MUS INBRED PWD were used. For LAB INBRED C57BL6/J, 15 two weeks old specimens, 16 four weeks old specimens, and17 six weeks old specimens were killed by decapitation and scanned. 16 eight weeks old specimens were scanned alive. For MUS INBRED PWD, 15 two weeks old specimens, 15 four weeks old specimens, and 16 six weeks old specimens were killed by decapitation and scanned. 18 eight weeks old specimens were scanned alive.

For the investigation of age effects in adult mice, the following LAB INBRED C57BL/6J mice were scanned alive: 16 eight weeks old animals, 10 ten weeks old animals, 10 twelve weeks old animals, 10 fourteen weeks old animals, 4 sixteen weeks old animals, 4 eighteen weeks old animals, 4 nineteen weeks old animals, 4 twenty weeks old animals, 4 twenty-one weeks old animals, and 5 twenty-three weeks old animals.

Museum specimens were prepared by the collectors and museum staff using either detergent solution or dermestid beetles, but the preparation method is usually not fully documented (A. Helfricht, Senckenberg museum, pers. comm.). Only adult specimens were included in the study, juveniles being identified by the small size of their skull and mandibles and low toothwear. Only right hemimandibles were considered, except in cases where only the left hemimandible of a specimen was undamaged.

Preparation artifacts and plasticity

For the preparation with detergent solution, mouse heads were boiled for 1hr in tap water, cooled for 1 hour and then incubated in a solution of 12g/l commercial household detergent in tap water at 37°C for 3 days. The hemimandibles were then manually prepared from the heads and allowed to dry. For the preparation with dermestid beetles, mouse heads were preprepared by removing the fur and the brain, and the rest of soft tissue was removed by the beetles. The skulls were then frozen to kill remaining beetles and larvae, then thawed and allowed to dry. To investigate the effects of preparation on shape, skulls of 15 LAB INBRED C57BL/10J were prepared following each protocol. In addition 20 animals of DOM IRAN AHVAZ were scanned alive and afterwards prepared using the detergent protocol.

Two approaches were used for the hard diet/soft diet comparison. In the first we used laboratory outbred *M. m. domesticus* derived from animals trapped in the Massif Central in France (Ihle et al. 2006). The hard diet animals, 17 offspring from four breeding pairs, were fed on normal rodent pellets (Altromin standard diet No.1324). The soft diet animals, 14 offspring from two breeding pairs, were fed on a powder diet which has the same nutritional composition as the pellets (Altromin standard diet No. 1321). To ensure health of the mothers and to reduce maternal effects of the treatment, the food in the soft diet cages was changed from pellets to powder only 2 weeks after birth of the experimental animals, approximately two weeks before weaning. All mice in the hard diet/soft diet experiment were scanned alive at 8-10 weeks of age. Our second comparison was between 19 wild-caught *M. m. domesticus* from Ahvaz, Iran (DOM IRAN AHVAZ), and 24 F1 offspring of 6 breeding pairs formed from 9 mice caught during the same sampling trip and raised in cages with standard food. All mandibles were scanned without without bone preparation.

Data acquisition

All mandibles were scanned with a micro-computer tomograph (microCT - VivaCT 40, Scanco, Bruettisellen, Switzerland). Whenever possible, the material was scanned "fresh", (either alive but anesthetized with Rompun/Ketamine, or in a fresh, frozen/thawed cadaver, or in an ethanol-preserved specimen). This condition applies to all but the museum specimens. The resolution of the scans depended on the material: 21 or 33 µm for prepared bone, 33µm for alcohol preserved/fresh specimens, and 38µm for living mice. In order to produce two-dimensional lateral views of the hemimandibles for geometric Morphometrics, the scans were oriented as follows: hemimandibles were outlined in the tomographic slices, and the bone was segmented using a visually determined threshold of 230 mg HA/ccm. From the triangulated surface of the three-dimensionally reconstructed hemimandible, the major spatial axis of the hemimandible were automatically determined as described in (Laib et al. 2002), and the 3D image datasets (both original gray scale dataset as well as segmented binary dataset) were then aligned in 3D with the direction of the major axis. From the aligned digital gray-scale dataset, a virtual 2D x-ray image was produced by adding the linear attenuation coefficients (the gray scale image values) in lateral direction (A. Laib, Scanco, Bruetisellen, pers. comm.). All these operations were done with the built-in software of the microCT system. Fourteen landmarks were digitized on each hemimandible for geometric Morphometrics using tpsDig2 (Rohlf 2005) and tpsUtil (Rohlf 2004), producing a set of 28 raw coordinates for each specimen (Fig.1).

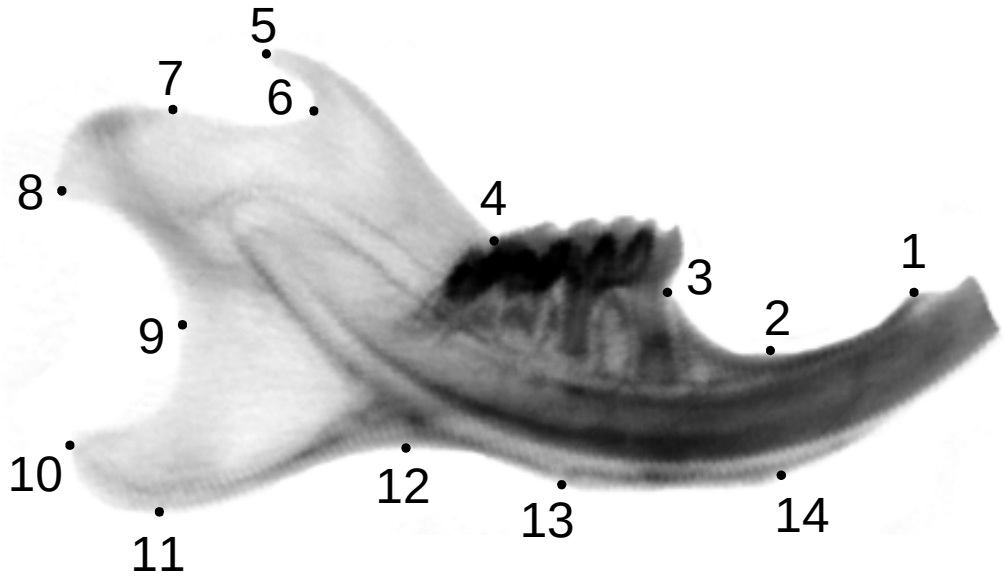


Fig.1. Landmarks on the outline of the mouse hemimandible used in this thesis.

Landmarks and digitization

The landmarks assessed in this study are depicted in Fig.1, anatomical descriptions are given in SUPPLEMENT 2. In order to estimate the measurement (= digitization) error, the whole reference dataset was digitized twice at different times and distances were compared between equivalent samples. The average difference between Mahalanobis distances from both digitizations was found to be 0.17, the maximum difference was 0.61 (4% and 14% of the average Mahalanobis distance in the reference dataset), and the correlation between both sets of distances is 0.97. In order to reduce the impact of this digitization error on the overall results, the average of both digitizations was used in the subsequent analyses.

Statistical analyses

All geometric morphometric analyses were performed in MorphoJ (Klingenberg 2008). Analyses were carried out on different subsets of the material. The raw coordinates of each subset were subjected to a Procrustes fit in MorphoJ, whereby variation due to position, orientation and size was removed from the data, leaving only shape variation for further analysis. CVA (canonical variates analysis) was used to calculate the Mahalanobis and Procrustes distances between all samples in a given dataset and to produce visualizations of the shape vectors associated with the CV axis. Discriminant function analysis was used as an additional tool for estimation of distinctness by numbers of misassigned specimens. PCA of Procrustes coordinates and correlation of PCA scores with centroid size were used to investigate potential effects of size, and regression of Procrustes coordinates on age was used to determine the effect of age on mandible shape in adult mice. Regression of Procrustes coordinates, reghression scores of Procrustes coordinates on dummy variables encoding group

membership and of cVA scores on trait scores was done in MorphoJ. Plots of CVA scores, Mahalanobis distances and amounts of variance explained were produced in R (R development core team 2008). T-test and ANOVA on categories of Procrustes and Mahalanobis distances were done in SPSS. Neighbour-Joining trees based on Mahalanobis distances were calculated in MEGA version 4 (Tamura et al. 2007).Calculation of interlandmark distances from raw coordinate data was done using the "Euclidean distance matrix (EDMA)" function in PAST (Hammer et al. 2001). ILMD difference vectors and trait scores were calculated in Excel.

Chapter 1: Technical issues and error sources

I. Procrustes vs. Mahalanobis distances

In order to decide whether to use Procrustes or Mahalanobis distances as a measure of biological shape disparity, I compared both measures with respect to their biological information content. This can be done by comparing their power to discriminate between intra- and interspecific distances. Procrustes and Mahalanobis distances for our full reference dataset are shown in Fig.2.

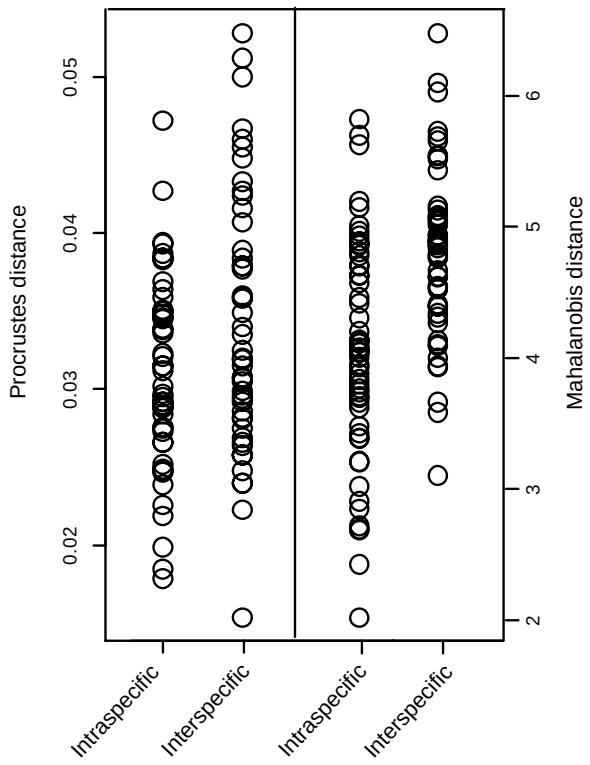


Fig. 2. Comparison of Procrustes and Mahalanobis distances with respect to discrimination between intra- and interspecific distances in our wild mouse reference dataset. For Procrustes distances, the means of the intra- and interspecific distances were 0.030 (s.d. = 0.007) and 0.034 (s.d. = 0.008), respectively ($p = 0.017$). For Mahalanobis distances, the means of the intra- and interspecific distances were 4.06 (s.d. = 0.84) and 4.80 (s.d. = 0.67), respectively ($p = 2.2\ 10^{-6}$).

We find that the distinction between intraspecific and interspecific distances is better with Mahalanobis distances than with Procrustes distances. Hence, since the Mahalanobis distances appear to be biologically more informative in our dataset, we used them for all further analyses.

II. Further technical issues

The analyses presented in this thesis rely on Mahalanobis distances, calculated from Procrustes-fitted landmark coordinates as well as on length measurements of interlandmark distances computed from landmark coordinates. I assume that the Mahalanobis distances are most sensitive to errors, because they accumulate the errors from the entire landmark configurations. Therefore I use the impact of error factors on Mahalanobis distances as a conservative estimator of introduced error. This impact is characterised using the Mahalanobis distances between samples known to be affected by a given error factor and control samples, and between these samples and the samples contained in the dataset for the wild mouse shape space. For each case, I will give the mean and the maximum absolute value of the difference between their respective distances to the samples within the reference dataset, and the correlation between the latter distances for the probe and the control. All correlations reported in this chapter are significant at $p < 0.001$ level.

1) Measurement error

In order to estimate measurement (=digitization) error, I digitized the reference dataset twice
The average difference between pairwise Mahalanobis distances for both digitizations is 0.17; the maximum difference is 0.61 (4% and 14% of the average Mahalanobis distance in the reference dataset), and the correlation between both sets of distances is 0.97.
I conclude that measurement error introduces some noise.

2) Orientation inside the scanner

In order to estimate the error introduced by different orientation of specimens inside the scanner, I scanned one sample of 16 specimens twice, each time in different random orientations.
The average difference between the Mahalanobis distance of the two resulting datasets to the samples in the reference dataset is 0.29, the maximum difference is 0.56 (7% and 13% of the average Mahalanobis distance in the reference dataset), and the correlation between both sets of distances is 0.98.

I conclude that orientation inside the scanner introduces no significant error additional to measurement error.

3) Preparation

In order to estimate the impact of preparation on Mahalanobis distances, I used 30 heads of C57BL/10J mice. These were scanned prior to preparation. 15 were subsequently prepared using the detergens protocol, and 15 were prepared by dermestid beetles. The prepared mandibles were scanned, the landmark data of all prepared mandibles were combined and compared to the data from the unprepared heads.
Furthermore, I prepared 17 heads of DOM LAB IRAN AHVAZ GEN1 mice, 11 of which had previously been scanned unprepared and were part of the Dom Lab Iran Ahvaz Gen1 dataset, using the detergent protocol and measured the error introduced by this. Note that 6 specimens were thus not originally included in the control (DOM LAB IRAN AHVAZ GEN1) dataset.
The average difference between the Mahalanobis distance of the two resulting datasets to the samples in the reference dataset is 0.65, the maximum difference is 1.04 (15% and 24% of the average Mahalanobis distance in the reference dataset), and the correlation between both sets of distances is 0.96. For the Iranian dataset, the results were: max. and average Mahalanobis distance difference = 0.62 and 0.29, respectively.
I conclude that preparation introduces some amount of error.

4) Sample size

With 14 landmarks, the minimal sample size according to recommendations in the literature (Mitteroecker and Gunz 2009) should be 28 (the number of dimensions times the number of landmarks). The majority of the samples used in this thesis is smaller than that (see material and methods). 9 of the 15 samples in the dataset for the wild mouse shape space consist of 20-30 specimens. I used these to empirically assess the effect of reducing sample size down to 15 specimens in a CVA. From each of these samples, I arbitrarily removed specimens until all samples consisted of 15 specimens. Then I combined the reduced samples with the unreduced smaller samples from the reference dataset and calculated the Mahalanobis distances. For the 86 distances involving reduced samples, I analyzed the differences to the distances from the unreduced dataset.

The average difference between the Mahalanobis distance of the two datasets is 0.28, the maximum difference is 0.96 (6% and 22% of the average Mahalanobis distance in the reference dataset), and the correlation between both sets of distances is 0.94.
I conclude that while reductions in sample size down to 15 specimens introduce error, the estimation of Mahalanobis distances is still robust enough to be useful.

5) Sampling errors related to size and age

Size

If there was a strong allometric vector related to size as a proxy for age in adult mice, and if this vector was the same regardless of genotype or population, then phenotypic differences between populations might simply reflect age differences between populations. In order to test this, I calculated PCA scores for the reference dataset and assessed the correlation of the scores for the first three axes with centroid size. None of the PCA axis was distinct, and the first three axes together explained 50.37 % of the total variance. The correlation of all three axes with centroid size was non-significant and close to zero.
I conclude that size does not determine shape in a general fashion in mandibles of adult mice.

Age

To assess the influence of adult age on shape, I regressed shape on age in a dataset of C57Bl76J mice in 10 age classes spanning 8 to 23 weeks of age (see material and methods). In this regression, age explained 9.54 % of total shape variance and 12.87 % of size variance.
I conclude that shape does change subtly with age.

Chapter 2: Wild mice

How large is the natural variation of mandible shape? In a comparative context, this means: how large is variation within populations compared to variation between populations, and how large is variation between conspecific populations with respect to variation between closely related species? An intuitively accessible answer to these questions can be given using the results of pairwise discriminant function analysis for all population pairs in our wild mouse dataset. Of 56 conspecific pairs, 7 (13%) are not significantly distinct at the 0.05 level, and 1 out of 49 cross-specific pairs is not significantly different at the 0.05 level. Discriminant function analysis gives the numbers of mis-assigned specimens between sample pairs after cross-validation. This is a measure of how well specimens can be assigned to the correct group based on shape, i.e. how distinct two samples are in shape. For conspecific pairs, on average 15% of specimens are mis-assigned, the maximum being at 41%. For cross-species pairs, on average 11% of specimens are mis-assigned, the maximum being 29%. Cross-species distinctness is thus on average somewhat larger than intraspecific distinctness. The difference between conspecific vs. cross-specific distinctness becomes clearer when one takes out the Kerguelen populations, which are, as shown below, dramatically distinct from all other *M. musculus* and which can be considered a special case in the face of their unusual history and habitat. Then, the average conspecific misassignment score becomes 20%, whereas 12% of specimens are misassigned on average in cross-specific pairs.

Beyond these generally indicative numbers, it is of interest to get an idea of the topology of the shape space, or the "network of similarities" of all the populations. There are three complementary ways to achieve a graphical representation of this.

The simplest, if incomplete, is by scatterplots of the most important canonical variates axis. Fig.3. shows scatterplots of the first four axes.

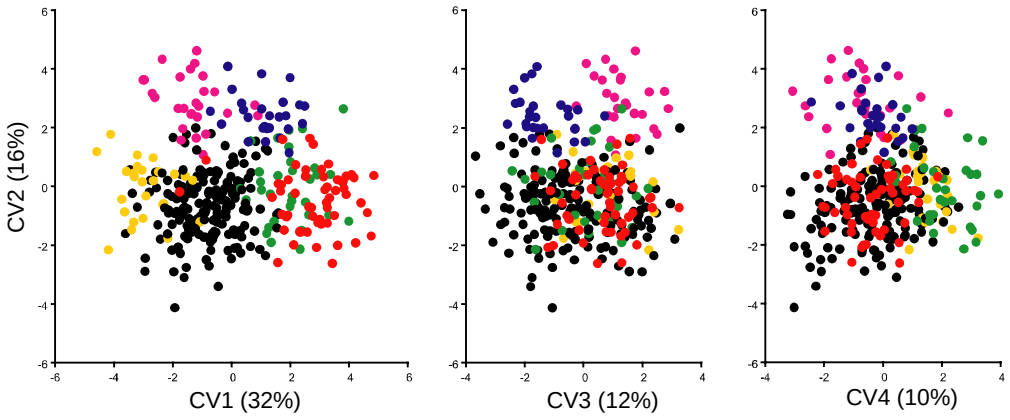


Fig.3. Scatterplots of the first four CV axis of the wild mouse reference dataset. Black: Mus musculus, except for: orange: Dom Spain PudeMont, pink: Dom Kerg Guillou, blue: Dom Kerg Cochons; red: Mus spretus, green: Mus macedonicus. In brackets: proportion of variance between groups explained.

Axis 1 is determined by the difference between *Mus musculus* and the other two species. DOM SPAIN PUDEMONT is special by being most distinct from the other two species on this axis, which is remarkable in so far as these specimens have been collected together with the *Mus spretus* sample SPR SPAIN PUDEMONT in a small village in the Pyrenees. In the museum drawer, these specimens were mixed within the same series, so they had apparently lived in close sympatry and had not been told apart by the collector. Axis 1 also separates DOM KERG GUILLOU and DOM KERG COCHONS from each other (see discussion). Axis 2 separates DOM KERG GUILLOU and DOM KERG COCHONS from the other groups, and axis 3 again distinguishes between DOM KERG GUILLOU and DOM KERG COCHONS. Axis 4 reflects the distinction between *Mus macedonicus* and *Mus spretus*.

While this plot can visualize in a simple way some major features of the shape space topology, it is limited in that only the differences between some groups can be highlighted without making the plot hard to read, and also the information on the higher CV axis is discarded.

A more complete way to represent the shape space is by a plot of Mahalanobis distances ordered in suitable categories. In Fig.4, the Mahalanobis distances of each category are shown in one column. Because all Mahalanobis distances are shown, this plot contains the complete information from the CVA. Each dot represents a Mahalanobis distance between two groups. Small distances – points near the bottom of the plot –mean that the groups have similar mandible shapes. The categories have been established by ordering the distances on a spreadsheet and searching manually for interesting patterns.

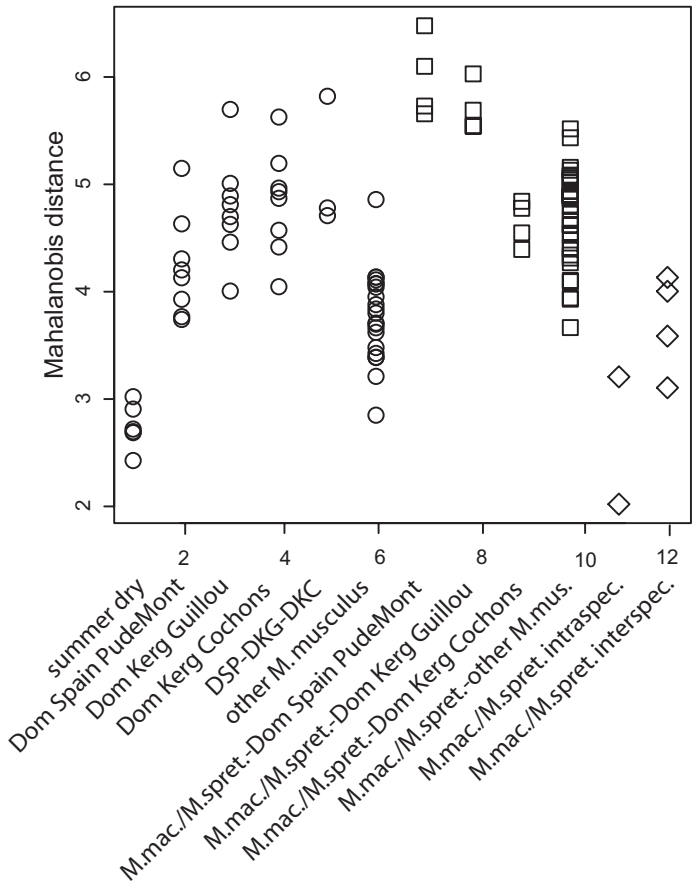


Fig.4. Mahalanobis distances in the wild mouse reference dataset. Circles: Intraspecific distances in *Mus musculus*, squares: interspecific distances between *Mus musculus* and *Mus macedonicus*/*Mus spretus*; diamonds: intra- and interspecific distances within and between *Mus macedonicus* and *Mus spretus*.

Column one shows that there is a cluster of populations – DOM EGYPT, DOM SICILY, DOM IRAN TEHERAN and DOM IRAN AHVAZ – which all have very similar mandible shapes. These samples are identical with the subset of *M. musculus* populations living in summer-dry regions according to the Koeppen-Geiger world climate map. At the same time, they may represent more ancient settlings of house mice than the other populations (see disussion).

Columns 2-5 show that there are three most divergent populations of *M. musculus*, DOM DOM SPAIN PUDEMONT, DOM KERG GUILLOU and DOM KERG COCHONS, which was visible already from the CVA plots. The remaining intraspecific distances in column 6 are intermediate. No major pattern was found in bulk of interspecific distances between *Mus musculus* and the *M. macedonicus*/*M. spretus* clade represented in column 10. A comparison of the distribution of distances in this column with columns 2-5 shows that the degree of divergence of DOM SPAIN PUDEMONT and the Kerguelen mice resembles a typical interspecific degree of divergence. Columns 7 and 8 show that DOM SPAIN PUDEMONT and DOM KERG GUILLOU are also highly divergent from *M. spretus* and *M. macedonicus*. DOM SPAIN PUDEMONT has diverged from *M. musculus* in an opposite direction in shape space

(Axis 1 in Fig. 3.). DOM KERG COCHONS, however, shown in column 9, seems to have become partially more similar to *M. macedonicus*/*M. spretus* by diverging from *M. musculus*, as also visible on axis 1 in Fig. 3. Columns 11 and 12, finally, reflect the smaller degree of intra- and interspecific divergence in *M. macedonicus* and *M. spretus* relative to *M. musculus*.

The scatterplot of Mahalanobis distances in categories has the advantage that it is easy to compare directly between distances, and it is the most direct way to show the complete information. However, the assignment of categories is an arbitrary compromise in the choice of which features to highlight, and a large part of the more detailed information becomes inaccessible because labelling all data points would again make the plot illegible.

The third way of representing the shape space is to construct a Neighbour-Joining tree from the pairwise matrix of Mahalanobis distances (Fig.5).

Fig. 5. Neighbour-Joining tree based on the pairwise matrix of Mahalanobis distances in the wild mouse reference dataset.

The most important features of the tree reflect the insights gained already from the other two representations. *M. macedonicus* and *M. spretus* sit on a common branch; DOM EGYPT, DOM SICILY, DOM IRAN TEHERAN and DOM IRAN AHVAZ form a cluster of closely related shapes;

DOM SPAIN PUDEMONT, DOM KERG GUILLOU and DOM KERG COCHONS have long branches, DOM KERG COCHONS being more closely related to *M. macedonicus* and *M. spretus* and DOM KERG GUILLOU to DOM SPAIN PUDEMONT. Additional features which appear in the tree are the closer relationship between the two samples from Germany, DOM GER FRANKFURT and DOM GER MUNICH (but note how large these distances still appear compared to the "summer-dry" cluster) and between the "eastern" populations, MUS HUNGARY and CAS JOHNSTON ATOLL.

Chapter 3: the role of phenotypic plasticity in shape differences

I assessed the amount of shape change introduced by environmental factors by quantifying the impact of two environmental "factors" on genetically similar mice: life in the wild vs. life in the laboratory, and, traditionally, hard diet vs. soft diet.

A direct comparison between different environments is between wild-caught animals and F1 offspring in the laboratory of wild-caught animals from the same natural population. This was done for the samples DOM IRAN AHVAZ (wild-caught) and DOM LAB IRAN AHVAZ GEN1 (lab-reared). A second approach to assess the influence of environmental differences is by feeding mice from the same population with different diets. This was done for DOM LAB FRANCE GEN3/4 mice and mice from the same outbred stock raised on soft diet.

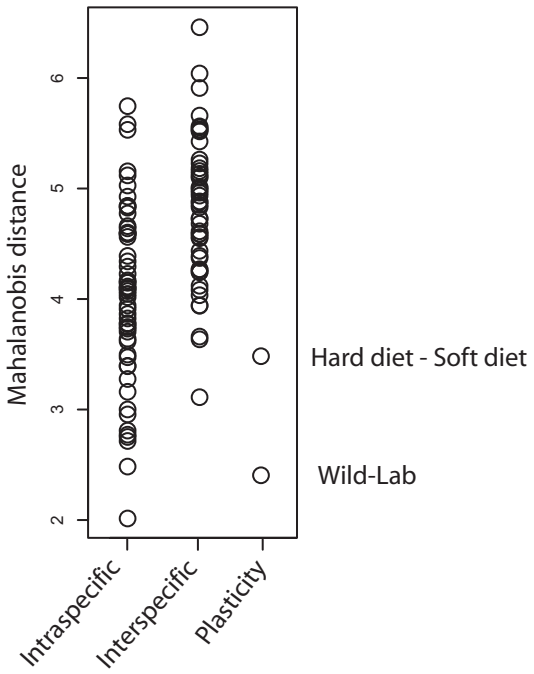


Fig. 6. Effects of environmental factors on mandible shape. Mahalanobis distances between hard-diet and soft-diet animals and between wild-caught animals and first-generation-cage-raised animals from the same population (Dom Iran Ahvaz) compared to intra-and interspecific distances in our wild mouse reference dataset.

Fig. 6. shows the Mahalanobis distances between "treatment" and "control" groups for both cases. The difference between DOM LAB IRAN AHVAZ GEN1 and DOM IRAN AHVAZ (Mahalanobis distance 2.4) is smaller than all others, except one intraspecific distance between wild-caught mice, although one can assume that both environments differ strongly from each other. The difference between hard diet and soft diet animals is considerably larger (Mahalanobis distance 3.5). It should be noted, however, that this experiment is plagued by

the problem that “control” and “treatment” come from different generations of the same stock (see material and methods), a factor which is likely to significantly affect mandible shape (see chapter 5), but was only discovered after the feeding experiment was finished. Hence, I expect that this distance would be smaller if it would be repeated under conditions that control for this additional factor.

Fig. 7. shows the shape difference between hard diet and soft diet animals. The shape change is mostly located in the angular process, which is more massively developed in hard diet animals. This is comparable with the shape changes found in similar feeding experiments by Renaud et al. (2010).

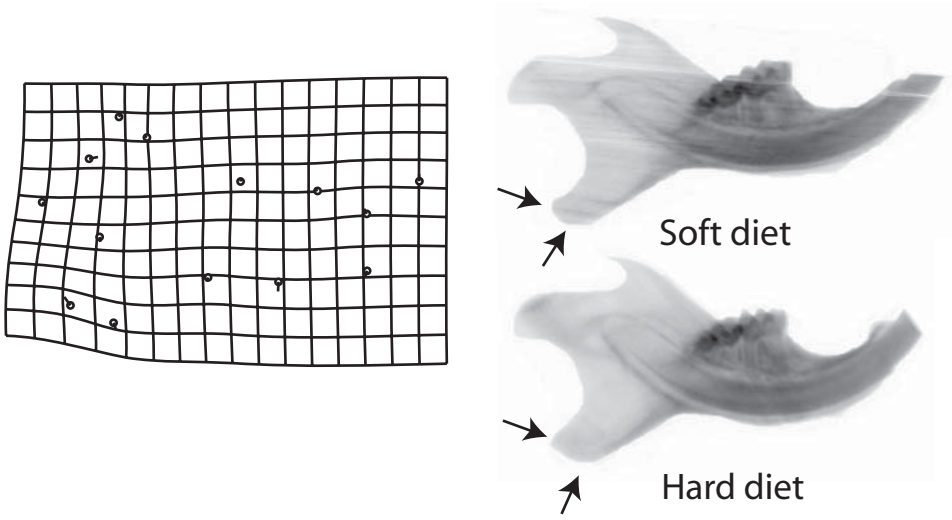


Fig.7. Mandibles of a soft diet and a hard diet animal. Shape change is located primarily in the angular process.

Chapter 4: divergence of inbred mice from wild mice

From a simple quantitative genetic perspective of additivity of genetic effects, one would not expect that inbreeding would drastically change the phenotype of an inbred line from the mean of the base population. After all, what happens during inbreeding is that some random combination of alleles from the base population is made homozygous. Some haplotype blocks containing recessive lethal alleles will be lost, but that should be all. As soon as an inbred line has overcome inbreeding depression and is happily reproducing, the phenotype should look normal.

Our initial motivation for starting a screening of inbred mouse strains was to identify strains which differ in mandible shape strongly enough to be suitable for genetic mapping. While the mapping project was eventually not continued, it turned out that the mandibles of most inbred mice look very different from those of wild mice. Fig. 8. shows what this looks like in a CVA plot.

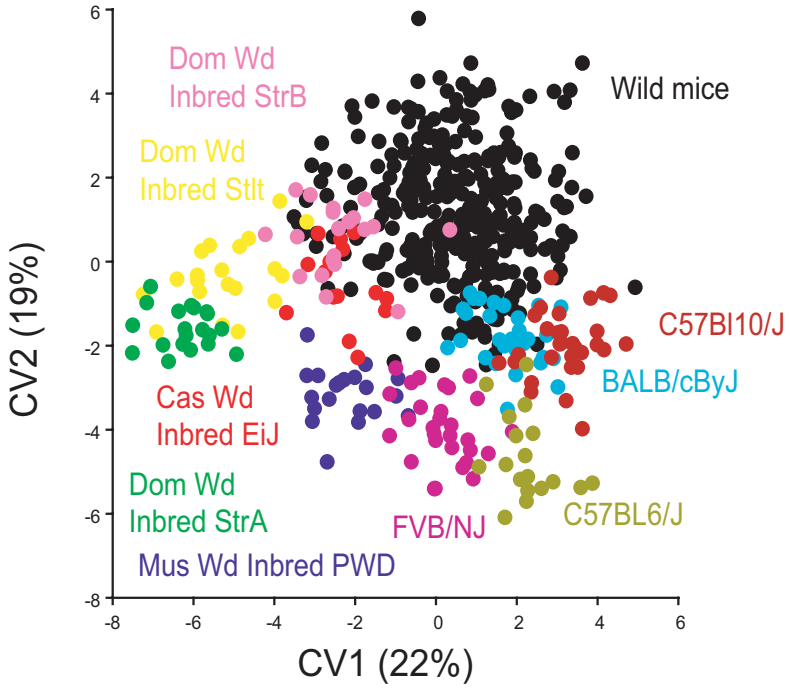


Fig. 8. Scatterplot of the first two CVs of the reference dataset together with nine inbred strains. All wild mice are shown in black, although individual wild mouse populations were coded as separate groups in the CVA. In brackets: proportion of variance between groups explained.

Three observations can be made on this plot: 1) in the coordinate system of the first two CV axes, which together explain about 40% of the total variance between groups, most inbred strains appear to lie outside or at least at the periphery of the wild shape space. 2) the

locations of the inbred strains are limited to a fraction of the periphery of the wild mouse shape space, i.e. the directions of deviation from the wild mice are biased. 3) the classical inbred strains, C57BL/10J, C57BL/6J, BALB/cByJ, and FVB/NJ, are located in a different region of the shape space than the wild-derived inbred strains, DOM WD INBRED STRA, DOM WD INBRED STRB, DOM WD INBRED STLT, CAS WD INBRED CAST/EIJ and MUS WD INBRED PWD.

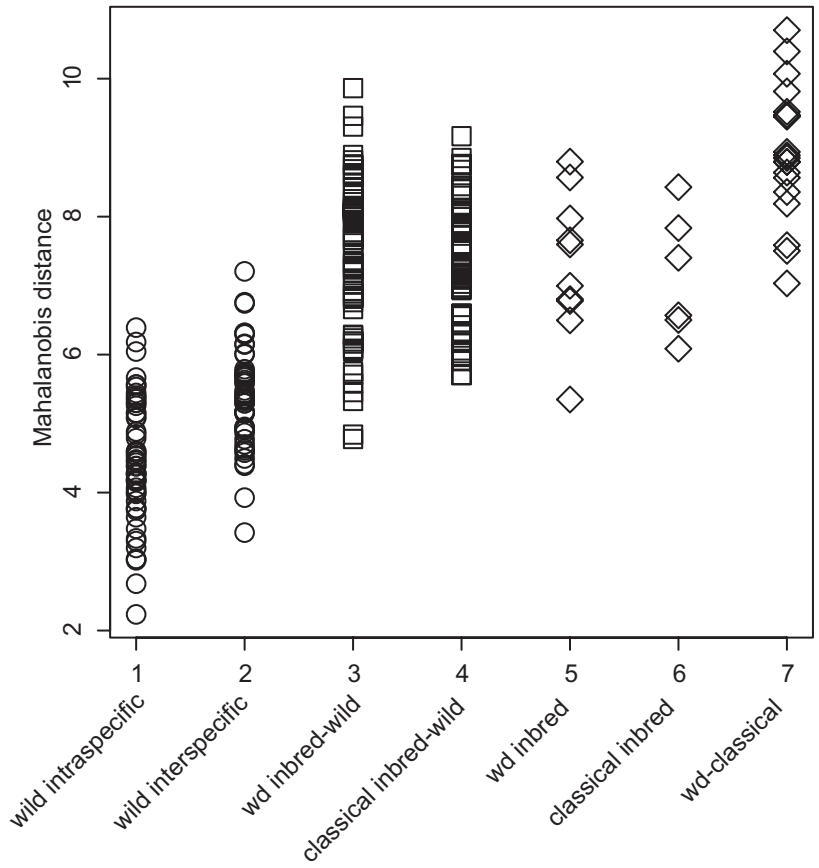


Fig. 9. Mahalanobis distances between wild and inbred mice. Circles: distances between wild mouse populations; squares: distances between inbred strains and wild mouse populations; diamonds: distances between inbred strains.

The distances are easier to compare in a scatterplot (Fig. 9.). It turns out that the distances between inbred mice and wild mice and among inbred strains are on average much larger than differences between the sister species in nature. Furthermore, the distances separating wild-derived and inbred strains are on average even larger, suggesting that these two categories of strains tend to occupy different regions in the shape space, as already visible in the CVA plot in Fig. 8.

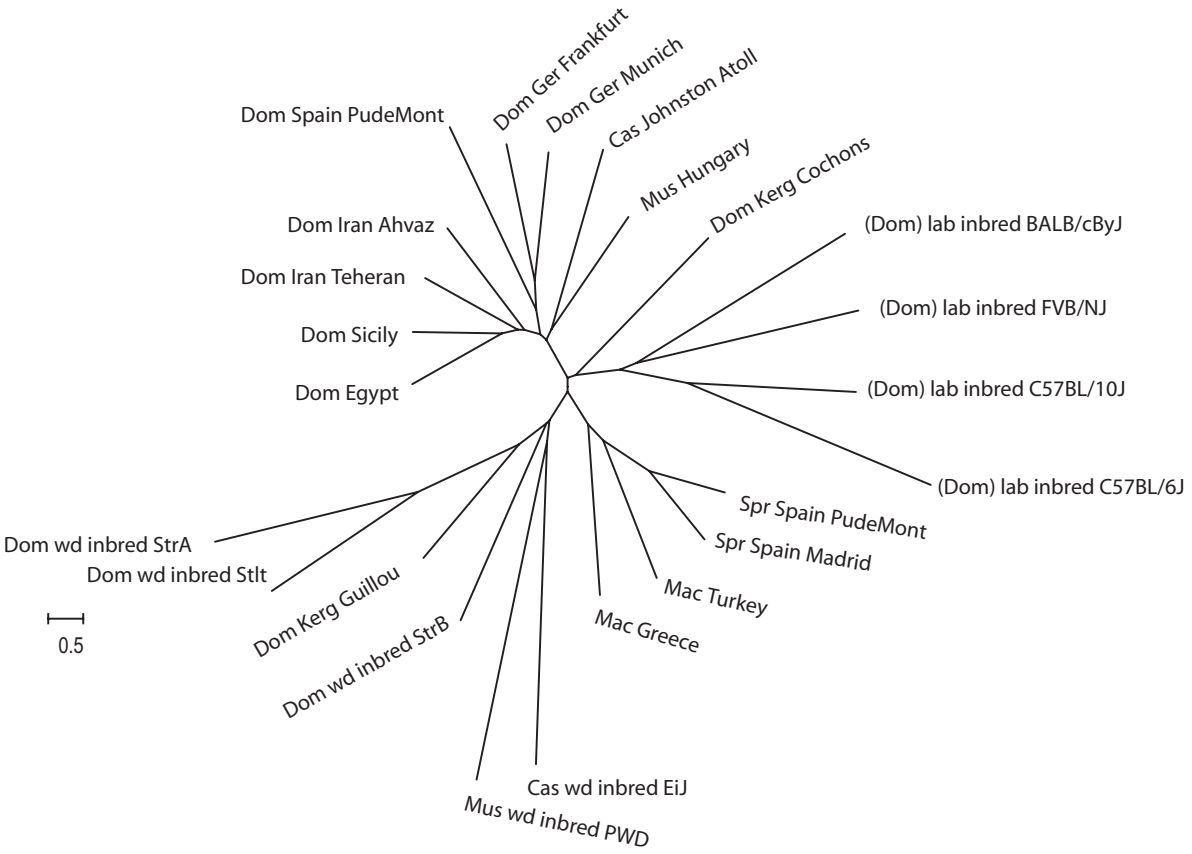


Fig.10. Neigbour-Joining tree based on the matrix of Mahalanobis distances between wild mouse populations and inbred strains.

Fig.10 summarizes the topology of the shape space as a Neighbour-Joining tree based on Mahalanobis distances. It can be observed that 1) each inbred strain is represented by a long branch, i.e. the mandible shape of each inbred mouse strain is highly distinct; 2) classical inbred strains form a separate clade; 3) the Kerguelen mice show affinities to the inbred strains, DOM KERG COCHONS to the classical ones and DOM KERG GUILLOU to the wild-derived ones.

Chapter 5: Wild-derived outbred mice

The strong divergence of inbred mice from wild mice may be a consequence of inbreeding, but during the course of the study, I found indications that even mice that are kept under outbreeding conditions in the laboratory show mandible shape changes over time. This was found for the samples of the outbred stocks which are being kept in the mousehouse in Plön: populations from Iran, France, Khazachstan, Germany (Cologne/Bonn), Spain, and Taiwan. These mice belong to different species and have been kept in the laboratory for a number of generations after having been initiated using wild-caught mice. In the case of the Iranian mice, the samples include wild-caught individuals from the base population (DOM IRAN AHVAZ), 1st generation (DOM LAB IRAN AHVAZ GEN1) and 3rd generation (DOM LAB IRAN AHVAZ GEN3) samples. Table 2. gives an overview:

Sample	Species	Generation
DOM LAB IRAN AHVAZ GEN1	*M. m. domesticus*	1
MUS LAB KHAZ GEN1	*M. m. musculus*	1
DOM LAB IRAN AHVAZ GEN3	*M. m. domesticus*	3
DOM LAB GER GEN3	*M. m. domesticus*	3
DOM LAB FRANCE GEN3/4	*M. m. domesticus*	3 and 4
DOM WD (INBRED) TAIWAN	*M. m. castaneus*	part. inbred
DOM WD (INBRED) SPR SPAIN MADRID	*Mus spretus*	part. inbred

Table 2. Outbred wild-derived populations kept in Plön.

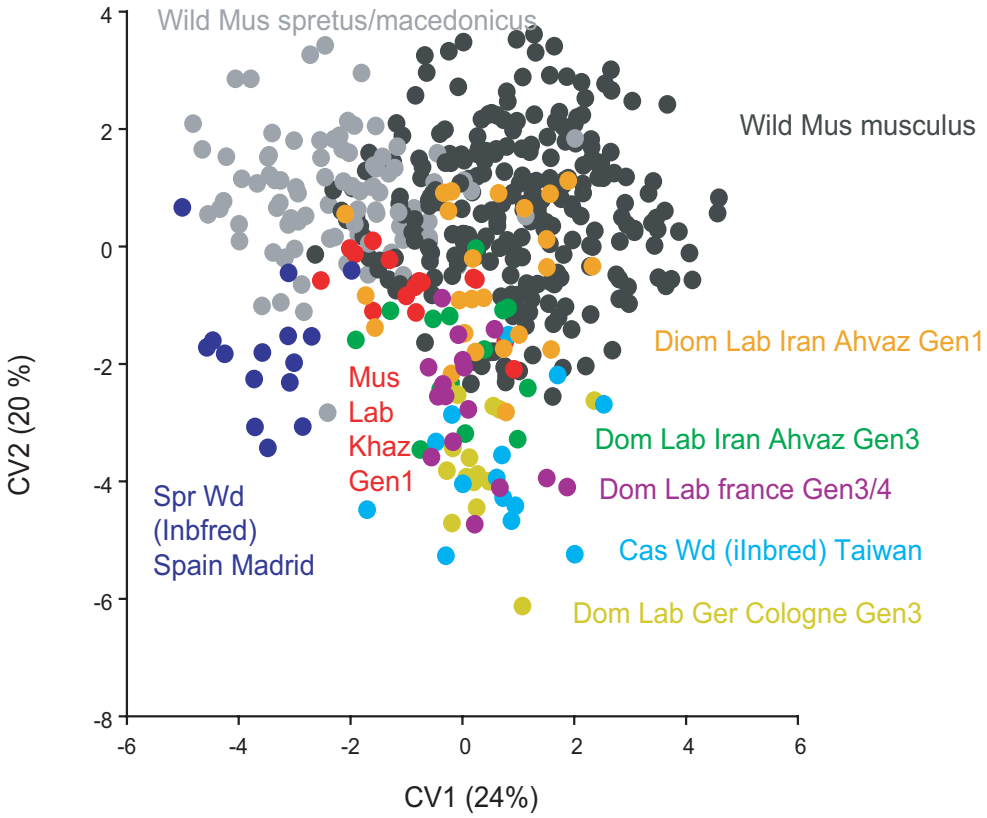


Fig.11. Scatterplot of the first two CV axes of wild mice and wild-derived outbred mouse populations.

In a scatterplot of the first two CV axes of these samples together with the wild mice (Fig. 11) it can be seen that the outbred mice lie mostly at the periphery or outside the shape space of the wild mice. The effect is smallest for the stocks which have been kept in the laboratory for only one generation, DOM LAB IRAN AHVAZ GEN1 and MUS LAB KHAZ GEN1. The stocks which have been bred in the laboratory for three or more Generations, lie mostly outside the wild shape space. Note especially the difference between DOM LAB IRAN AHVAZ GEN1 and DOM LAB IRAN AHVAZ GEN3, which are derived from the same wild-caught mice (DOM IRAN AHVAZ) and are only separated by two generations of breeding in the laboratory. Similarly as for the inbred mice, the outbred mice seem to "leave" the wild shape space in similar directions. Note also that the difference between *Mus spretus* and *Mus musculus* (CV1) is conserved in the course of this transition.

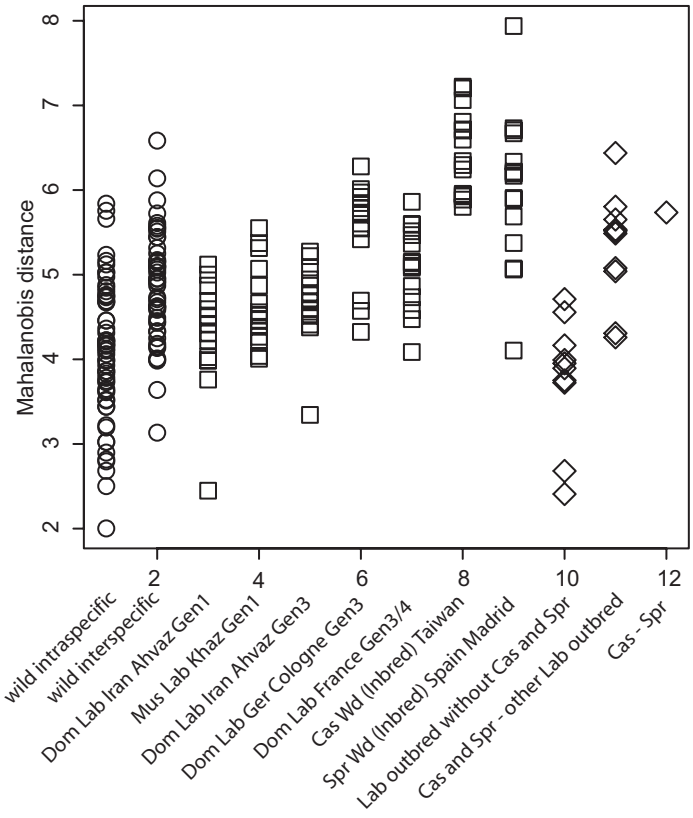


Fig.12.Mahalanobis distances grouped into categories.

Fig.12 shows the Mahalanobis distances between outbred and wild mice. Because it is not immediately visually clear whether the distances between the samples of outbred mice and wild mice are more similar to intraspecific or to interspecific distances, I tested these differences using a Bonferroni-post-hoc corrected ANOVA. The results are shown in Table 3.

Sample	Diff. from intraspec (p-value)	Diff. from interspec (p-value)
DOM LAB IRAN AHVAZ GEN1	1	0
MUS LAB KHAZ GEN1	0.291	0.224
DOM LAB IRAN AHVAZ GEN3	0.225	1
DOM LAB GER GEN3	0	0.062
DOM LAB FRANCE GEN3/4	0	1
DOM WD (INBRED) THAI	0	0
DOM WD (INBRED) SPR SPAIN MADRID	0	0

Table 3. Results of an ANOVA testing for differences between Mahalanobis distances of wild-derived outbred populations to wild mice and wild intra-and interspecific Mahalanobis distances.

It turns out that the distances between first generation outbred mice (DOM LAB IRAN AHVAZ GEN1 and MUS LAB KHAZ GEN1) are more similar to intraspecific distances between wild-

caught mice, whereas the distances of outbred mice in more advanced generations of lab breeding to wild-caught mice resemble interspecific distances between wild-caught mice.
The distances of the partially inbred mice, DOM WD (INBRED) TAIWAN and DOM WD (INBRED) SPR SPAIN MADRID, to wild mice are on average even larger than typical interspecific distances between wild mice. The distances among those outbred stocks which are not partially inbred resemble typical intraspecific distances, which is congruent with the observation from the CVA plot that they seem to have diverged from the wild mice in similar directions.

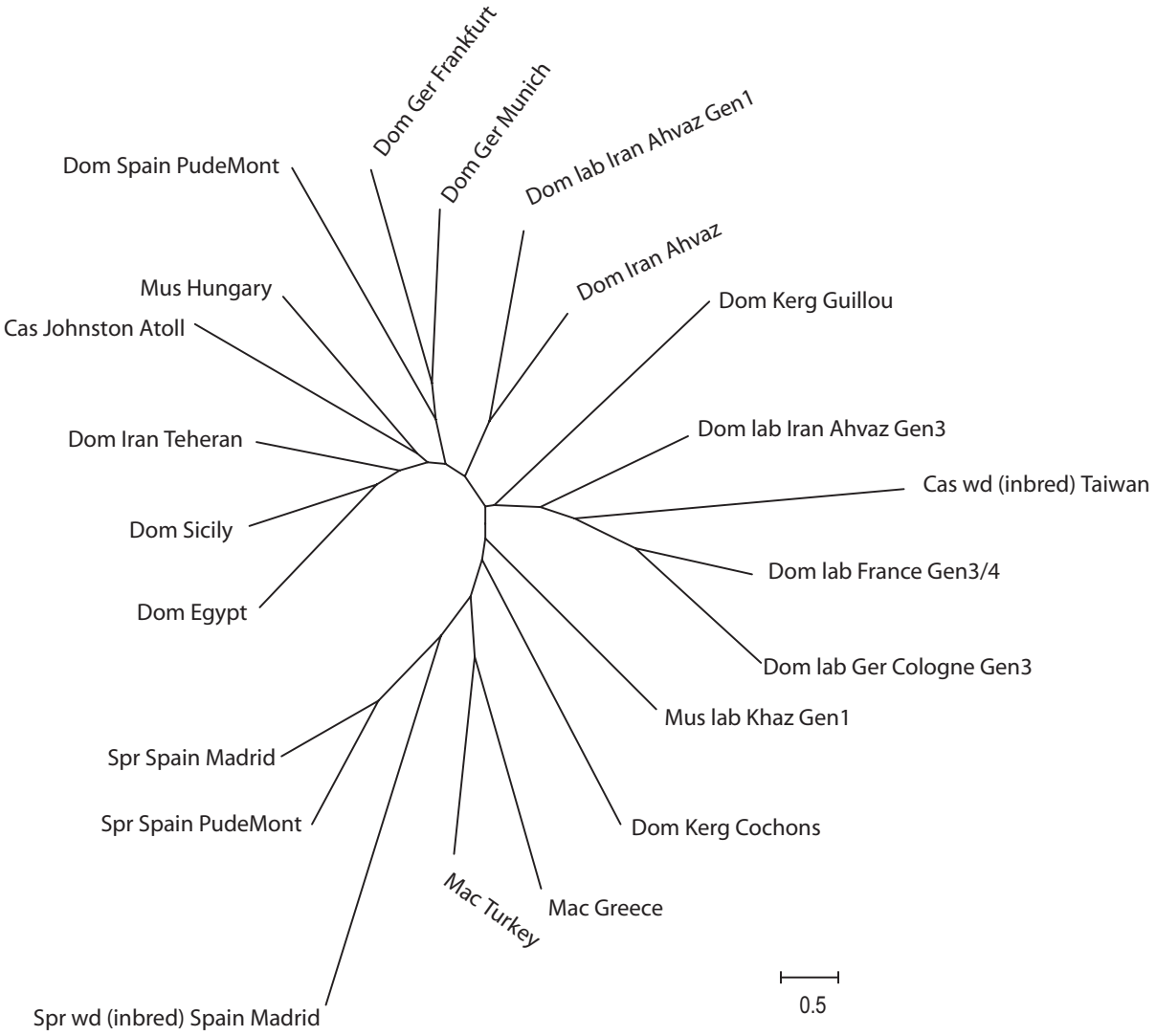


Fig.13. Neighbour-Joining tree based on the Matrix of mahalanobis distances between wild mouse populations and wild-derived outbred populations.

Fig.13 shows a Neighbour-Joining tree of wild and outbred mice based on Mahalanobis distances. DOM LAB IRAN AHVAZ GEN3, DOM WD (INBRED) TAIWAN , DOM LAB FRANCE GEN3/4 and DOM LAB GER GEN3 form a common branch, supporting the interpretation that they have diverged in a common direction from wild mice, leading to distances between them

and their wild conspecifics which are rather in the range of interspecific distances. Note the separation between DOM LAB IRAN AHVAZ GEN3 and the DOM LAB IRAN AHVAZ GEN1/DOM IRAN AHVAZ branch thus introduced.

Chapter 6: The effects of outcrossing on shape in the F1 respective to the parental strains/populations and to wild mice

If inbreeding does change mandible shape because it makes certain deleterious alleles homozygous or because it reduces the allelic diversity within individuals, then this change should be at least partially or wholly reversible by outcrossing. In order to assess this hypothesis, I analyzed F1 offspring of outcrosses between C57BL6/J and CAST/EIJ, between C57BL/6J and PWD, and between Cast/EiJ and PWD, respectively. The CVA results on this outcross triangle are shown in Fig.14.

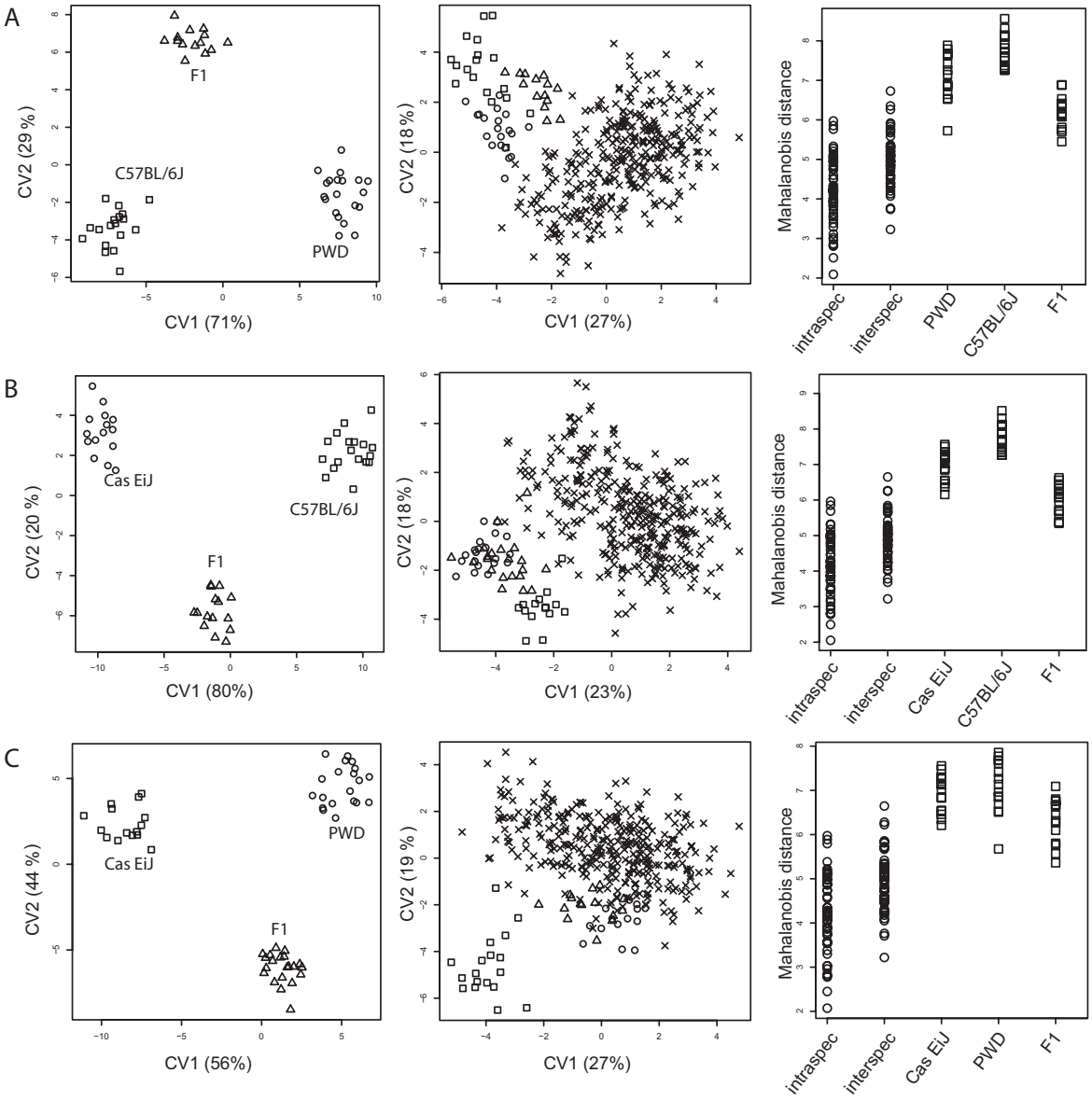


Fig.14. Shape space changes in F1 crosses between inbred mouse strains. A) C57BL/6J x PWD, B) C57BL/6J x CAST/EIJ, C) PWD x CAST/ EIJ. The left panels represent scatterplots of the first two CVs of a CVA including only the strains with their F1 offspring, the middle panels represent the same CVA,

but including the wild mouse populations (crosses) and the right panels show Mahalanobis distances grouped into categories. Distance comparisons of strains and F1 are against wild mice.

Two observations can be made on the outcrosses: 1) in all three crosses, the F1 are more or less intermediate between the parental strains on CVA1, but deviate from the intermediate shape along CVA2 in the leftmost panels in Fig.14, i. e. there are strong "dominance deviations"; 2) in all three crosses, the F1 are more similar to wild mice than either parental strain, as partially visible on the middle panels and clearly visible on the rightmost panels in Fig.14. However, it is important to note that this "reversal" to a more "wildtype" shape by outcrossing is not complete; the distances between each F1 and wild mice are still very large.

These results raised interest in the question whether the strong "dominance deviations" observed in outcrosses of inbred mice would also occur in outcrosses of outbred mice. Therefore I analyzed the F1 offspring of an outcross between our wild-derived outbred stocks DOM LAB IRAN AHVAZ and DOM LAB FRANCE. Note, however, that the parentals of this outcross were in different generations: the DOM LAB IRAN AHVAZ parentals were in generation1 (phenotypically not very divergent from wild mice), and the DOM LAB FRANCE parentals were in generation 2, and thus probably phenotypically less divergent from wild mice than DOM LAB FRANCE GEN3/4. Unfortunately, heads of French animals of generation 2 have not been scanned.

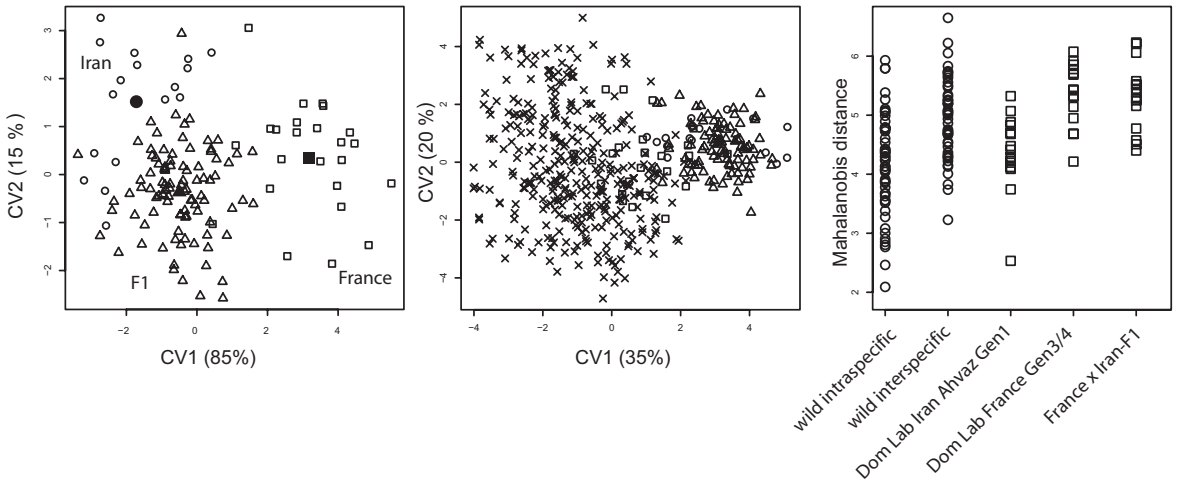


Fig.15. Shape space changes in an F1 cross between DOM LAB IRAN AHVAZ and DOM LAB FRANCE. The left panel represents a scatterplot of the first two CVs of a CVA including only the populations with their F1 offspring (group means marked as filled symbols), the middle panel represents the same CVA, but including the wild mouse populations (crosses) and the right panels show Mahalanobis distances grouped into categories. Distance comparisons of populations and F1 are against wild mice.

Fig.15 shows that "dominance deviations" do also occur in outcrosses between outbred populations. Generally, variances are larger in outbred mice and the shape differences are smaller than between inbred strains, therefore the effect appears less drastic. However, the

group mean of the outcross F1 is clearly off the line connecting the parental populations. However, a phenotypic “reversal” to a more “wildtype”-like shape is not observed in this case.

Discussion of chapters 1-6 and introduction to chapters 7 and 8

Here I discuss the results from chapters 1-6. From this discussion, a motivation for the investigations presented in chapters 7 and 8 is derived. These chapters are introduced through a connecting passage.

I Discussion of chapters 1-6

1. Technical issues in comparisons among heterogeneous samples

Before I could test for the influence of potential confounding factors in my dataset, the issue of the most suitable measure of disparity between populations had to be settled. In Geometric Morphometrics, Procrustes distances are the Euclidean distances between group means in the tangent space, whereas in the calculation of Mahalanobis distances, the data are corrected for common within-group covariance resulting in a distortion of the original data space (Klingenberg and Monteiro 2005, Zelditch et al. 2005). There is currently no general agreement on whether Mahalanobis distances are suitable for biological interpretation (Mitteroecker and Gunz 2009), and they are not interpretable in terms of geometric relationships between shapes in the original dataspace. However, if one is interested in disparity between groups, they may be more informative because they are not influenced by similarity of within-group covariances among groups. Despite cautionary comments in the literature (Mitteroecker and Gunz 2009), Mahalanobis distances turned out to be a better discriminator of the biological disparity associated with species differences than Procrustes distances (Fig.2). This argument is sufficient to make them preferable in the search for biologically relevant shape differences.

Strictly technical problems consist of low sample sizes (unavoidable when using museum material and sometimes also for laboratory mice due to limitations of breeding facilities and money as well as considerations of animal welfare), potential influence of different preparation and sample conservation methods on shape, orientation of specimens inside the scanner, and digitization error. I have tested all of the technical factors and found that they introduce different amounts of error in the estimation of Mahalanobis distances.

Measurement errors set the upper limit of resolution in the data. Fortunately, their overall effect on Mahalanobis distances is negligible. Different orientation of specimens inside the scanner does not introduce any significant additional error.

When using CVA based on landmark data, the general recommendation found in the literature is not to use sample sizes of less specimens per group than the number of landmarks times the number of dimensions, i.e. degrees of freedom in the analysis (Mitteroecker and Gunz 2009). In theory, one would otherwise have to reduce the number of landmarks, discarding valuable information. I use 14 landmarks to describe mandible shape. Thus, according to the recommendation in the literature, my minimum sample size should be 28 specimens. Fortunately, it turns out that the effect of reducing sample size from 20-30 down to 15 is not all that dramatic. As for the Procrustes vs. Mahalanobis distance controversy, empirical assessment of the effects of specific factors does not necessarily bear out the theoretical/statistical concerns. In the end, I decided to include samples with at least 14 specimens in the analysis.

Museum samples do mostly consist of prepared bones. There are two common methods of preparing the skull: boiling in water and subsequent incubation in detergent solution in order to degrade the soft tissues (Davis 1983), and preparation by exposition to dermestid beetles. For museum specimens, it is usually not documented which method has been used. In my assessment of both preparation methods, it appears that they have tangible effects on shape and on Mahalanobis distances. The effect is, however, no worse than the effect of small sample sizes.

In summary, the estimation of generally reliable Mahalanobis distances seems possible in spite of the limitations imposed by the available samples and methods. The caveat is of course that it is unknown how these sources of noise add up. There is however no reason to assume that this causes a systematic bias. Also, the fact that genetically coherent groups, such as the two populations from Germany (DOM GER MUNICH and DOM GER FRANKFURT) as well as the laboratory inbred strains (all four LAB INBRED) form clades in the distance tree (Fig.10), supports the notion that the shape assessments reflect a genetic basis and is not dominated by noise.

Further error may be introduced as a result of sampling error with respect to sex, size and age if these factors have a strong enough influence on shape. Sexual dimorphism for mandible shape is reported to be absent in wild mice by (Davis 1983). The PCA assessment of size in the context of the large wild mouse dataset reported here is the first of its kind and demonstrates strikingly that allometry vectors are at best very population-specific, in which

case accounting for size would be useless. The influence of age is more difficult to test, as it inevitably requires large numbers of mice of known age, spanning a wide range of ages, and then it is still unlikely that the age-shape trajectory across the shape space can be generalized across populations. My preliminary assessment of age in 2 to 5 months old mice suggests that age explains only a small fraction of within-population variance. The only references which I could find in the literature about the average lifespan of a wild mouse indicate that it is about 3 months (Berry and Jakobson 1971, Bomford 1987). Most of the laboratory mice analysed in this thesis were between 2 and 3 months old when their heads were scanned. Set aside the impossibility to account for age in a general way if age-shape trajectories are population-specific, I conclude that it is unlikely that age differences are a major problem in comparisons of shape in the context of my data. For these reasons, I ignored size, age and sex as error sources and as shape determinants in my analyses.

2. Plasticity vs. genetic differences

It is known that genetically identical organisms can develop in different ways as a specific reaction to environmental differences. On the other hand, genetically different organisms may develop differently in identical environments. Because environment and genes do interact, it is misleading to ask whether any phenotypic difference is caused by “genes or environment”. One can, however, ask whether environmental differences are sufficient to explain observed phenotypic variation, or whether genetic differences must be assumed to account for it.

In the specific context given here, I ask whether the shape changes intoduced by environmental factors are possibly large enough to account for shape differences as found between conspecific wild mouse populations, and whether there is evidence for genetically based shape differences between the populations, irrespective of environmental factors.

To assess the influence of environmental factors on mice of otherwise comparable genetic constitution, I used a field-laboratory comparison and a traditional hard diet-soft diet experiment. The latter scenario has been investigated several times (e.g., Levrini et al. 2003, Maki et al. 2002, Yamada and Kimmel 1991, Renaud and Auffray 2009, Renaud et al. 2010). However, the magnitude of the shape change introduced by the difference in diet consistency in relation to the shape differences between natural populations has not been addressed in these studies.

Here, I find that the environmental factors do have an effect of mandible shape. In the case of the field-laboratory comparison, this difference is almost negligibly small compared to the

wild conspecific interpopulation differences. The difference introduced by the diet consistency is larger, although still much smaller than the average wild conspecific interpopulation difference (Fig.6). There are two reasons to assume that this datapoint is an overestimation of mandible shape changes introduced by diet consistency differences in nature. The first is simply that house mice are generalists which are not prone to dietary specialization. The second is that our experiment is slightly flawed insofar as there is a one to two generation difference between treatment and control (see material and methods). In chapter 5 of the results I show that mandible shape changes as a function of the number of generations in captivity (at least for the first few generations). This effect is expected to increase the disparity between treatment and control beyond the difference introduced by diet consistency.

At the same time, evidence for genetic differences causing shape differences between conspecific populations comes from the wild-derived outbred house mouse populations kept in the laboratory in Plön. The differences among these populations, which are kept under identical conditions, are on average no smaller than the wild conspecific interpopulation differences (Fig.12). In summary, the available evidence suggests that shape differences between conspecific populations in nature are mostly caused by differences in genetic composition irrespective of environmental factors.

3. Natural selection on mandible shape

In the present context, stabilizing selection and positive selection have to be considered as potentially important processes for mandible shape in general and for specific populations.

In order to infer positive selection from purely phenotypic data, departures of specific populations from a presumed ancestral phenotypic state are required. The broad sampling of house mice and sister species in this thesis allows the construction of a shape space, in the context of which particularly divergent populations can be identified. Since the populations from the Mediterranean and the Near East representing the supposedly oldest settlements of house mice in the entire dataset are not particularly divergent, it can be assumed that the average shape resembles the ancestral shape of the mandible and that the divergent populations are thus evolutionarily derived. There are three populations for which this seems to be the case: DOM SPAIN PUDEMONT, DOM KERG GUILLOU and DOM KERG COCHONS (Fig.3-5). DOM SPAIN PUDEMONT is differentiated from other house mice along CV1 in Fig.3, i.e. in the opposite direction of the *Mus musculus –Mus spretus* / *Mus macedonicus*

difference. This population lived in close sympatry with a *Mus spretus* population in a village in the Spanish Pyrenees (the samples in the museum drawer were mixed and universally labelled "*Mus musculus*"). One might therefore think that this divergence might represent a case of character displacement (a special case of positive selection). However, it must be cautioned that it is hard to see why this character displacement should cause an exactly opposed shape difference in the mandible with respect to wild mice. Besides this, they are geographically in sympatry, but their habitat is thought to be very different, as *Mus musculus* lives in houses and barns, *Mus spretus* in open fields, so they are not necessarily expected to be in competition with each other. However, character displacement is the only ad hoc hypothesis which comes to mind to explain the derived phenotypic status of DOM SPAIN PUDEMONT.

The situation is different for DOM KERG GUILLOU and DOM KERG COCHONS. These populations have originated from two independent late 19th century colonization events of the subantarctic Kerguelen archipelago as a consequence of human transportation (Hardouin, in prep.). The environment on these islands can be considered as very different from Western European environments the mice came from. Mean temperatures are below 10°C all year round (Le Roux et al. 2002) with high precipitations and much wind, and the islands are characterised as "polar tundra" on the Koeppen-Geiger climate map (http://koeppen-geiger.vu-wien.ac.at/index.htm). Le Roux et al. (2002) found that earthworms are an important part of the diet of mice on Guillou, which might not be eaten so much by mice in other habitats, and which might influence mandible shape via adaptive evolution or phenotypic plasticity. At the same time, predators are absent on the islands. In Fig.3. it can be seen that both Kerguelen populations have diverged partially in the same direction from other wild mice, namely along CV2. It must be cautioned, however, that they are at the same time very divergent from each other along CV1 and along CV3, and that they are globally rather distinct from each other (Fig.4, 5). This divergence and also their large divergence from other wild house mice may be attributed to a bottleneck through which they have gone in the course of the colonization events (see below). Nevertheless, the common aspect of divergence along CV2 might indicate some degree of parallel adaptations to the Kerguelen environment. But this evidence is at best somewhat suggestive.

The second type of natural selection that may have acted on mandible shape in wild mice is stabilizing selection. I postpone the question of general stabilizing selection restricting the total shape space of wild mice to the second more general part of the discussion. At present, I

have to consider the case of the populations from the Mediterranean and the Near East, DOM EGYPT, DOM SICILY, DOM IRAN AHVAZ and DOM IRAN TEHERAN. As can be seen in Fig.4 and 5, these four populations form an exclusive cluster of closely related shapes. At the same time, they do exclusively share the "privilege" to live in summer-dry regions according to the Koeppen-Geiger map of world climate (http://koeppen-geiger.vu-wien.ac.at/index.htm). One might thus be tempted to speculate that life under summer-dry conditions imposes special requirements on mandible shape leading to these similarities via stabilizing selection on the ancestral phenotype. The populations from Iran represent the oldest settlements of house mice in my dataset (based on molecular and fossil evidence, it has been shown that *M. m. domesticus* mice have spread from Iran into the Near East and Northern Africa and not more than 3,000 years ago into Sicily and Western Europe (Cucchi et al. 2005)). This idea entails that the more distinct shapes of the Western and northern European house mice have diverged in different directions from the ancestral phenotype. As an alternative to adaptive radiation, neutral evolution and genealogical effects cannot be excluded. It is hard to see in which respect the more recent, not summer-dry colonization areas in Western and Northern Europe may represent a higher diversity of habitats than summer-dry habitats and the Near East, requiring adaptive radiation of mandible shape. An adaptive radiation of the house mice in summer-wet areas with effects on mandible shape, either by selection on shape itself or by correlated, pleiotropic effects of selection on other characters is a possible explanation, requiring further corroboration. A potential alternative could be a release of stabilizing selection in conjunction with epigenetic effects caused by subsampling from the ancestral allelic diversity in the course of colonization.

4. Divergence of inbred mice from wild mice and partial reversal by outcrossing

Two classes of inbred mice have to be considered. The so-called classical inbred strains have been established from admixed laboratory populations from various sources, containing genetic material from the three subspecies of the house mouse, *M. m. domesticus*, *M. m. musculus* and *M. m. castaneus* (Frazer et al. 2007). Wild-derived strains, on the other hand, have been established by brother-sister mating of offspring of wild-caught mice from specific populations. In order to understand the phenotypic outcomes of the inbreeding process, it is important to be clear about its population genetic implications. Inbreeding means on the one hand very strong purifying selection against lethal recessive alleles and positive selection on domestication traits such as reduced aggression, early reproduction and high fertility under

laboratory conditions etc (pleiotropic effects on the mandible during artificial selection on body composition have been documented by Atchley et al. (1990)). On the other hand, it implies a very strong population bottleneck. Thus, recombination cannot establish linkage equilibrium between selected alleles and the reminder of the genome, such that the selection process itself imposes strong drift on most of the genome. With respect to the phenotype, one has thus to consider the (pleiotropic) effects of the selected alleles and the general phenotypic consequences of strong drift, plus the general reduction in allelic diversity and heterozygosity.

In all of this, it is important to remember that the inbred mice do genetically represent a sample of wildtype alleles. Inbreeding is a rapid process, thus the occurrence of new mutations should be negligible. However, although selection and drift changes the genetic composition of the mice, one would not expect a strong genetically based phenotypic change as a consequence of inbreeding under a strictly additive model of genetic architecture.

The results of my phenotypic screen of inbred mouse strains do not conform to this expectation. To the contrary, not only are there huge phenotypic differences between wild and inbred mice, but the inbred strains themselves are much more different from each other than wild populations (Fig.8-10). The only way to resolve this conundrum is to assume a prevalence of non-additive genetic effects in the determination and inheritance of mandible shape. This conforms to our everyday experience: children do not look like perfect intermediates between their parents.

Under a non-additive model of genetic architecture including allele-specific interactions, it is easy to see that a general reduction of allelic diversity within an animal can have drastic phenotypic consequences: as the number of allele-specific interactions is reduced, the particular effects of the remaining alleles and interactions may become manifest instead of being cancelled out by other alleles and interactions.

Non-additive interactions of alleles at QTL loci have been shown to exist for the mouse mandible by Cheverud et al. 2004 and for the mouse skull by Burgio et al. 2009.

It is surprising that this effect, which gives the opportunity to assess the importance of non-additive genetic effects, has not received more attention in research and in the literature.

Lacy and Horner (1996) have studied the effect of inbreeding on skeletal development in *Rattus villosisimus* and found that it alters skull shape. There has also been a lot of interest in understanding the role of epistasis and bottlenecks in releasing new variation in the context of speciation models. Cheverud and Routman (1996) and Cheverud et al. (1999) have specifically addressed this for mice and found significant support for the epistasis model. Similarly, Bryant and Meffert (1990) found in bottleneck studies in house fly populations

(*Musca domestica*) that strong bottlenecks can indeed lead to a shift in shape space in comparison to wildtype variability.

To invoke non-additive genetic effects to explain the divergence of inbred from wild mice raises the question whether these effects are primary due to allele-unspecific effects (as for the "complementation" of "deleterious", "recessive" alleles or to allele-specific effects. If one assumes that allele-unspecific complemetation is restricted to within-locus interactions, the question becomes equivalent to distinguishing between dominance and epistasis.

It can be addressed by outcrossing inbred strains to each other. If allele-unspecific interaction plays a primary role, then outcrossing should lead to a more or less complete reversal of the F1 offspring to the wild phenotype because the large majority of "recessive" alleles will be complemented. If, on the other hand, allele-specific effects play a dominant role, then it is much more unlikely that these can be complemented by crossing together unrelated inbred strains. I analyzed the outcross triangle between C57BL/6J, Cast/EIJ and PWD. As a result, I find that there is indeed a reversal effect in the F1, i.e. they are phenotypically more similar to wild mice than either parent for all three outcrosses (Fig.14). This result is in itself an argument for epistasis, because in an additive model the F1 should not deviate from the intermediate shape between the parents (in fact, they are intermediate for some traits, see CV1 in the leftmost panels in Fig.14). The fact that this "epistatic deviation" brings the F1 closer to wild mice does signify, strictly speaking, that the divergence of inbred mice from wild mice is partly due to a lack of complementation, and that the outcrossing has reintroduced some of this complementation. If, as outlined complementation is primarily assumed to be relevant for within-locus interactions, then this result means that non-additive within-locus interactions play a significant role in shape determination. In this case, further non-complementary epistatic interactions between loci have to be assumed, because the phenotypic reversal is only partial (the Mahalanobis distances between F1 and wild mice are still large in Fig.14).

In summary, it appears that the divergence of inbred mice from wild mice and the partial reversal of this divergence by outcrossing make a strong case for the prevalence and importance of non-additive genetic interactions for the determination of an individuals phenotype. The place of these findings in the long-standing debate between adherents and opponents of the "additive world view" and their relevance for future research will be discussed in the synthetic part of the discussion.

5. Wild-derived outbred laboratory populations diverge from wild mice.

I found in the direct comparison between wild-caught animals and lab-raised F1 of wild-caught animals from the same population that that the within-generation phenotypic plastic response to the different environmental conditions affects shape only to a small degree. However, most wild-derived outbred populations kept in the laboratory have mandible shapes which are divergent from the wild shape space (Fig.11-13). At the same time, it turns out that the degree of this divergence is somewhat dependent on the number of generations of breeding in the laboratory (Fig.14, Table 3). Because the genetic composition of these animals is supposedly not different from wild mice, and within-generation plasticity has been found to be low, this is a surprise. Transgenerational phenotypic change in outbred captive mammals has been already reported by McPhee (2004). McPhee speculates that a release of constraint by stabilizing selection might be the reason for the phenotypic change. However, this hypothesis would just predict an increase in intrapopulational variance, not a change in mean shape. Additionally, it cannot explain why the phenotypic change leads to divergence from wild mice in a *similar direction* in shape space for all captive populations (Fig.11, 13), see also below). Another hypothesis would be selection for reproduction under laboratory conditions, but, as discussed above, unless the very alleles conferring this success have strong pleiotropic effects on mandible shape, the concomitant change in allelic composition of the population would tend to increase divergence also between the outbred populations in comparison to the wild shape space. Note that the phenotypic differences between the outbred populations are not larger than average conspecific interpopulation differences in nature (Fig.12). The phenotypic change observed does rather resemble a parallelism in the strict sense of independent parallel changes in all populations (Fig.R.11, 13; this observation is consolidated by an explicit assessment of "parallelisms" in results chapter 7, see below). The only "explanation" for this finding is to assume a potentially cumulative, transgenerational, epigenetically inherited environmental effect. Such an effect could be mediated by hormones and epigenetic mechanisms of inheritance such as DNA methylation. In the literature, there is a discussion about similar effects in connection with "stress" caused by extreme environments (Badyaev 2005). The problem with this "explanation" is, obviously, that there is not yet any evidence for epigenetic changes at the molecular level in captivity, and that evenif there was, the link between these changes and the mandible phenotype would remain unclear.

II Introduction to Chapters 7 and 8

Chapters 1-6 have dealt with determining distances in the natural morphospace of the mouse mandible and the inferences that we can derive from this knowledge on the underlying genetic architecture. The next two chapters will now deal with the next level of analysis, namely to infer from the patterns that I have observed specific growth processes (directions in the morphospace) that underlie the trait divergence and the corresponding evolutionary implications. The methods developed here are in some way ad hoc and exploratory, i.e. they go into directions that have to my knowledge not been explored in morphometrics before. Hence, it will be necessary to start with a few definitions and clarifications of terms to be used here.

In the CVA plots in Fig.8 and 11, inbred strains and wild-derived outbred populations digress into distinctly restricted segments of morphospace outside the wild mouse shape space. A similar pattern applies for the Kerguelen samples in Fig.3.
I will henceforth refer to these populations collectively as “derived populations”, and I will term the phenomenon of specificity or non-randomness of the directions of divergence as “parallelisms”. I use the latter term in the absence of a better designation, yet I will keep it in quotes in order to distinguish it from the idea of parallel adaptation to which the term refers most often in the literature and also because these “parallelisms” are not parallel in the sense of identical directions; instead the directions are just related to varying degrees. It must be borne in mind that the term “parallelism” is normally used in superficial, qualitative contexts rather than quantitative and multivariate. A specific term to denounce similarity of directions in multivariate data does not exist so far.

Because morphological structures are generated by structured growth processes (the „developmental choreography“), it is not expected that variation is generated at random with equal probability and to the same extent in all directions of the shape space. Insofar as this is the case, it is not unexpected that the directions of divergence of the Kerguelen populations (DOM KERG GUILLOU and DOM KERG COCHONS), the inbred mice and the wild-derived outbred populations from wild mice seem to be concentrated in certain segments of the shape space instead of showing a spherical distribution around the wild shape space.
The finding or the assumption that the production of variation by developmental processes is “biased” is often associated in the literature with the concepts of canalization or constraint (Salazar-Ciudad 2006a, 2007). I find both concepts difficult to use in the present context. Canalization (Waddington 1942) refers to the idea that a certain phenotype is produced in

spite of developmental or genetic disturbances of the developmental system. This does clearly not apply here. Constraint, on the other hand, invokes a relationship between evolution and development in the sense that certain directions of evolutionary change are impossible because the necessary variation cannot be produced due to limitations – “constraints”- of the developmental system (Maynard-Smith et al. 1985). However, none of my data suggest such a constraint.

On a purely descriptive level, my finding is that in specific contexts, variation is not spherical in the mathematical sense. I consider this *per se* as trivial for morphological structures. The nontrivial part is *which* are the most important directions of variation, not that there are such directions. I will try to tease apart these directions into their fundamental components (components in a biological rather than in a mathematical sense) in chapter 7. I call these components “major traits” or “traits”.

Concerning the methodology, the superficial observations from CVA plots which are intended to show different things in the first place seemed unsatisfying, because they do not allow quantification of the degree of relatedness of the directions of divergence, and because it is impossible to learn from them which features of the mandible are mainly involved in this phenomenon. The only tools which geometric morphometrics has to offer to compare directions of shape changes are graphical representations under the form of thin plate spline deformation grids. Whereas such visual comparisons are again qualitative, it is possible to calculate the angles between the vectors of landmark loadings represented in these plots. However, it is unclear how to interpret such angles, especially in the absence of a clear null hypothesis of unrelatedness – how small does an angle have to be in order to indicate relatedness of shape change? Therefore, in chapter 7, I develop a protocol based on distance measures, namley distances between landmarks or "inter landmark distance" (ILMD) to quantitatively compare directions of shape differences and to identify those traits which are conferring similarity between shape changes.

Chapter 7 – identification of traits in the divergence of inbred and captive mice and the mice from the Kerguelen islands

1) Demonstration of "parallelisms"

In order to circumvent the lack of geometric morphometric methods for comparing directions in shape space (see discussion), I decided to use interlandmark distances (ILMDs), which correspond to the length measurements of "traditional morphometrics" (i.e. statistics of length measurements before the advent of geometric morphometrics, Zelditch et al. (2004)).
The idea is simple: if two shapes differ in a similar way from a third shape, their differences from this shape should correlate.
It has already been mentioned in chapters 4 and 5 that the directions of divergence of inbred and wild-derived outbred mice (the "derived populations") from wild mice appear to be not randomly scattered around the wild shape space; instead, they seem to be concentrated in specific segments of the space, as seen in the CVA plots in Fig.4.a. and 5.a. and in the NJ trees in Fig.4.c. and 5.c. In a preliminary fashion, I entitle this phenomenon as "parallelisms"
In order to explicitly investigate the "parallelisms" whose presence is guessed from the geometric morphometric analyses, interlandmark distances were used. I looked for parallelisms between "derived populations" with respect to their "taxonomic base groups". Two "taxonomic base groups" were assumed: 1), *Mus musculus* (DOM EGYPT, DOM GER FRANKFURT, DOM IRAN AHVAZ, DOM IRAN TEHERAN, DOM GER MUNICH, DOM SPAIN PUDEMONT, DOM SICILY, MUS HUNGARY, CAS JOHNSTON ATOLL); 2), *Mus spretus* (SPRE SPAIN PUDEMONT, SPRE SPAIN MADRID). DOM KERG GUILLOU and DOM KERG COCHONS are not included in the *Mus musculus* base group, because they are considered as derived groups. The following "derived groups" (samples which are highly disparate from wild mice) were considered (Table 4):

Sample	Base group	Derived "by"
DOM KERG GUILLOU	*Mus musculus*	Kerguelen, inbreeding
DOM KERG COCHONS	*Mus musculus*	Kerguelen, inbreeding
DOM LAB IRAN AHVAZ GEN1	*Mus musculus*	lab breeding
DOM LAB FRANCE GEN3/4	*Mus musculus*	lab breeding
DOM WD INBRED STLT	*Mus musculus*	lab inbreeding
DOM WD INBRED STRA	*Mus musculus*	lab inbreeding
DOM WD INBRED STRB	*Mus musculus*	lab inbreeding
CAS WD INBRED CAST/EIJ	*Mus musculus*	lab inbreeding
MUS WD INBRED PWD	*Mus musculus*	lab inbreeding
BALB/CBYJ	*Mus musculus*	mixed genomes, lab inbreeding

FVB/NJ	*Mus musculus*	mixed genomes, lab inbreeding
C57BL/10J	*Mus musculus*	mixed genomes, lab inbreeding
C57BL/6J	*Mus musculus*	mixed genomes, lab inbreeding
FRANCE X IRAN F1	*Mus musculus*	lab breeding, outcrossing
C57BL/6J X CAST/EIJ F1	*Mus musculus*	lab inbreeding, outcrossing
C57BL/6J X PWD F1	*Mus musculus*	lab inbreeding, outcrossing
MUS LAB KHAZ	*Mus musculus*	lab breeding
DOM LAB IRAN AHVAZ GEN3	*Mus musculus*	lab breeding
CAS WD (INBRED) TAI	*Mus musculus*	lab breeding
SPRE WD (INBRED) SPAIN MADRID	*Mus spretus*	lab breeding/inbreeding
DOM LAB GER COLOGNE	*Mus musculus*	lab breeding/inbreeding

Table 4. Sample names, taxonomic base groups and method of "derivation" of the "derived populations" investigated in this thesis (see text).

The basic idea is to use the correlations between the vectors of differences of derived groups from base groups as a measure of parallelism.

The protocol for this is as follows:

1) calculate absolute ILMDs for all mice from landmark coordinates (91 distances per individual from 14 landmarks)
2) calculate the average ILMD vector (a vector of 91 elements) of each sample.
3) subtract the ILMD vector of the base group from the ILMD vector of each of its derived groups
4) scale the resulting ILMD difference (ILMDd) vectors by dividing them through the base group ILMD vector
5) calculate the correlations between the resulting vectors.

The resulting correlation matrix (SUPPLEMENT 7) can be transformed into a distance matrix by subtracting it from 1 and then be used to visualize the patterns of parallelism in a "Parallelism-Neighbour-Joining-Tree".

50% of the correlations (104 of 210) are significant at 0.001, 13% (27 of 210) are significant at 0.05, and 38% (79 of 210) are not significant. This means that captive, inbred and Kerguelen mice have diverged in related directions from wild mice more often than not.

In order to test the robustness of these correlations, I bootstrapped the ILMDs 10,000 times in R (script shown in SUPPLEMENT 5) For each of the 210 pairwise correlations between the derived populations, 91 ILMDs were sampled with replacement for each bootstrap. The bootstrap mean was always very close to the measured correlation: the average and the maximal absolute difference between them were 0.018 and 0.082, respectively. The average and maximal standard deviation of the bootstrapped correlation coefficients were 0.15 and

0.35, respectively. I thus assume that the observed correlations are robust and not due to just a few outlier ILMDs.

2) Description of the patterns of parallelism

Fig.16 shows a NJ-tree calculated from this correlation matrix after turning the correlations into distances by subtracting them from 1. The topology of the tree is mostly incongruent with "phylogenetic" relationships (for instance, populations derived from *M. m. domesticus* are scattered all over the topology instead of clustering together). As expected, the topology is mostly consistent with the shape space topology based on Mahalanobis distances as shown in Fig.10 and 13.

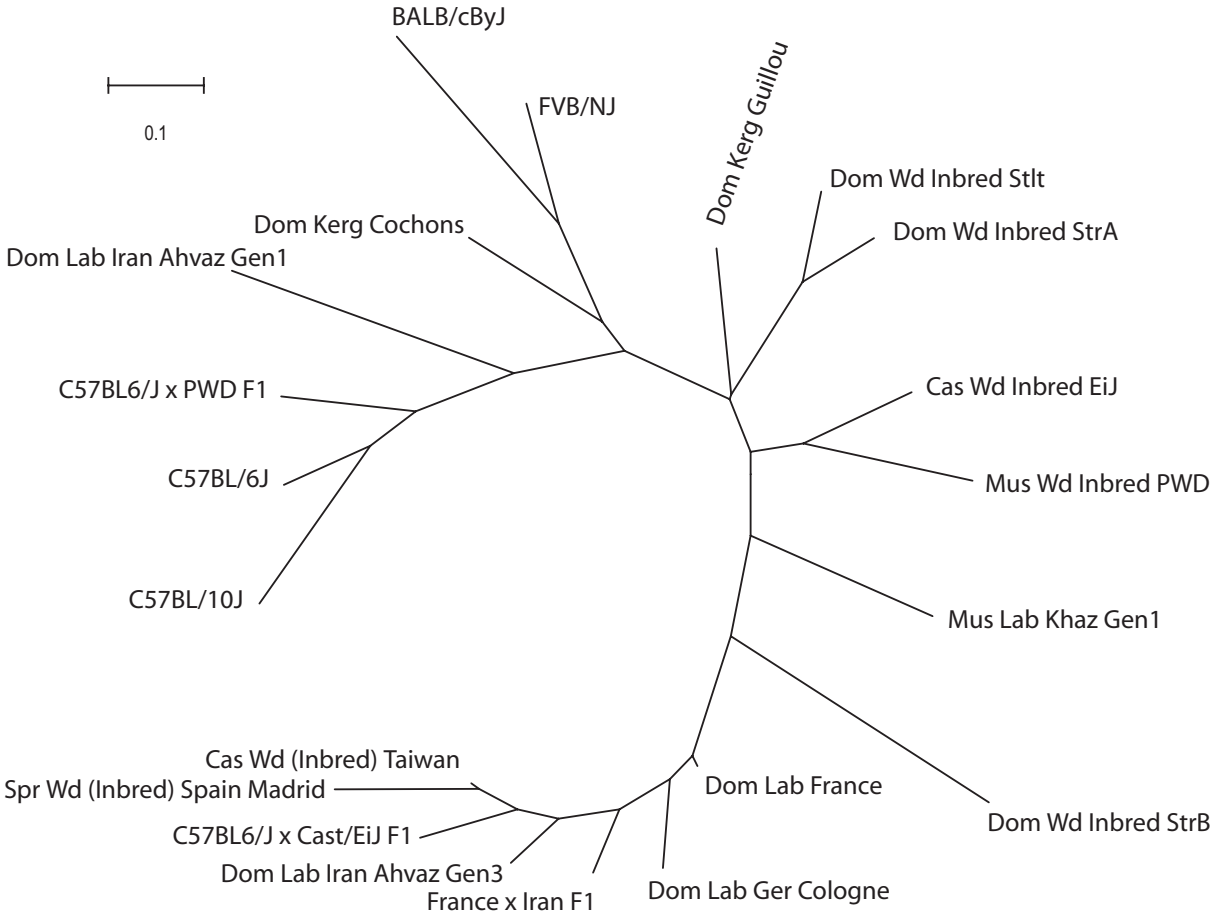


Fig.16. Neighbour-Joining tree based on the correlation matrix of ILMD difference vectors containing the information about how a derived population differs from its base population (the correlation matrix was transformed into a distance matrix by subtraction from 1 to calculate the tree) .

Looking at the distribution of the populations over the tree, it appears that outbred mouse populations share other features with respect to wild mice than inbred mouse populations. There are only three exceptions: DOM LAB IRAN AHVAZ GEN1 comes to lie on the side of the

inbred mice (on a long branch without any closer affinities). C57BL/6J x PWD F1, although an outcross, is placed near its parental strain C57BL/6J. DOM WD INBRED STRB comes to lie on the side of the outbred mice, again with a long branch.

3) A hypothesis of "parallelisms" in terms of constraints on extreme shape variation and discussion of its use in discovering the most variable aspects of shape in the mouse mandible

Before proceeding with the dissection of parallelisms between derived mouse populations, it is necessary to give a conceptual motivation for why and how this is done. I will argue that these populations represent an unique opportunity to identify developmentally defined major traits whose variation represents a significant amount of the total shape variation and which can be of heuristic value as informed hypotheses in developmental and genetic research. The line of argument is as follows:

- The number of measurable traits on the mouse mandible is in principle infinite.
- Not all dimensions can vary independently, i.e. variation cannot be isometric with respect to all traits. There are two sorts of constraints:
- 1) geometric constraints. Geometry imposes correlations among some traits.
- 2) developmental constraints. As a consequence of the supposedly hierarchical organization of the developmental processes, there is a limited number of major growth processes which define the general features of mandible shape (there may also be "minor" processes which locally refine shape). Each of these major growth processes has a specific effect on mandible shape. Because the number of these processes is limited and their effects are specific, they cannot make use of all geometrically available degrees of freedom. From the perspective of geometry, this (hierarchical) way of generating shape imposes developmental constraints.
- I assume further that each of these growth processes can be expressed to a higher or lesser degree, introducing axes of correlated variation. These directions of variation can be called "developmental traits".
- Developmental traits may affect overlapping regions of the mandible. Trade-offs between the expression of different traits may thus arise.
- In a genetically variable population, the expression of epistatically controlled traits will be balanced by epistasis (see discussion) to produce the average mandible shape. No trait will be extremely expressed.

- Inbreeding and other disturbances of the genetic and developmental system may lead to extreme expression of epistatically controlled traits. Mandible shape becomes thus “dysbalanced”.
- Through multiple independent instances of disturbance, it is possible to manifest the “dysbalanced shape space”. I hypothesize that this has been unconsciously done by the breeding and inbreeding of captive mice.
- The limitation of the number of major traits to be disturbed entails “parallelisms” in the "dysbalanced" shape space.
- The structure of the "dysbalanced" shape space can be used to identify the major expressed developmental traits which are normally almost “invisible” because of their balanced expression.

4) A rationale for the identification of developmental traits in ILMD differences

According to the above argumentation, “dysbalanced”, i.e. “overexpressed” developmental traits can be identified based on the ILMD differences between base and derived populations. How can this be done? Assuming that not all 91 ILMDs are independent traits, the task is to group together those which behave in a correlated manner because they are affected by the same hypothetical trait.

Since this amounts to a data reduction approach, conventionally one would use some ordination technique such as PCA to group ILMDs into independent factors explaining significant amounts of variance. The problems with PCA in the present situation are, however, twofold: first, the assumption of orthogonality of the factor axes is inappropriate here, because traits may be assumed to depend on each other in a hierarchical fashion and thus correlate positively or to be negatively correlated as a result of conflicting effects on shape in the case of extreme expression. Second, some ILMDs may be extremely variable between populations. These outliers could attract the PCA axes into wrong directions.

Another idea that might occur is to visualize the correlation structure of the ILMD or ILMDd (ILMD difference vectors, see above) directly using transformation into distances and calculation of a NJ-tree. However, the complexity of the relationships between 91 variables prevents any reasonable insights from the resulting topology.

Accordingly, I decided to analyse the ILMD dataset manually. I designed a procedure to reduce in a stepwise manner the number of potential traits and to discover the relationships

between the remaining ones. The method is necessarily crude, but sufficiently flexible to do justice to the biological reality. The analysis proceeds in the following way:

1) For each derived population, sort the ILMD difference vector, keep the 10 largest positive and the 10 largest negative ILMD differences and discard the remaining ones. The analysis becomes thus restricted to the most significant shape differences between base and derived populations. In the following steps, positive and negative ILMDd are considered independently, i.e. an instance where the distance between LM5 and LM6 has become larger in a derived population compared to the base population is considered unrelated to instances where this ILMD has become smaller. This is necessary, because the deviations of ILMDs in the derived populations with respect to the base populations may not be symmetric, and the correlation structures may be different for decreased and increased ILMDs.
2) Among the remaining ILMDd, discard those which occur less than four times in the total dataset remaining after step1. (Only those ILMD differences which are potentially involved in parallelisms are kept.)
3) In a pairwise matrix of the remaining ILMDd, enter in each cell the number of times the corresponding pair of ILMDd occurs together in a derived population.
4) On the resulting diagonally symmetrical matrix, apply the following criterion of "correlation": if, for a given pair of ILMDd A and B, it is true that B co-occurs with A more than half the number of times that A occurs in total, and if the same applies vice versa, then this pair of ILMDd is assumed to be "correlated", otherwise not (a R script for counting the co-occurences is shown in SUPPLEMENT 6).
5) For each ILMDd, write a list containing those ILMDd with which it is correlated. From the resulting list of lists, remove those ILMDd lists which are redundant because they are a subset of another ILMDd list, and those which do not contain at least one negative and one positive ILMDd (a positive and a negative ILMDd are required to calculate a scale-free score, see below).
6) For each of the remaining sublists, calculate a score from the original population-wise averaged ILMDs (not the ILMDd) of the derived populations by dividing the sum over all populations of those ILMDs whose ILMDd is positive through the sum over all populations of those ILMDs whose ILMDd is negative in the list. Each score is given the name of the ILMDd which "leads" the sublist.

7) Compile a matrix of all scores, where the columns are the scores and the rows are the populations.
8) Calculate the correlation matrix between the scores and transform it into a distance matrix by subtracting it from 1. Then calculate a NJ tree from the distance matrix. Use the resulting topology to infer the relations between the "leading" ILMDd and to further reduce the number of traits.

5) Results of the stepwise trait discovery procedure

Step1:

For 21 derived populations, I retained the 10 largest positive and the 10 largest negative ILMDd. Thus a dataset of 420 ILMDd was retained (SUPPLEMENT 8).

53 of 91 ILMDs were represented among the 10 largest positive ILMDd in the derived populations. 50 of 91 ILMDs were represented among the 10 largest negative ILMDd in the derived populations. Only 30 ILMDs were represented in both sets, indicating a high degree of asymmetry of positive and negative ILMDd. 18 of 91 ILMDs were not at all represented in both sets.

Step2:

Here, I discarded all ILMDd which occurred in less than four derived populations among the ILMDd retained in step 1. 24 of 53 positive ILMDd occurred at least 4 times. 21 of 50 negative ILMDd occurred at least 4 times. Together, the "frequent" ILMDd made up 79% (331) of the 420 ILMDd retained in step 1. While 39 of 91 ILMDs were still represented, only 6 of them were represented both as positive and as negative ILMDd among the frequent ILMDd. Asymmetry of positive and negative directions is thus very pronounced here. In summary, somewhat less than half of the 91 ILMDs in the derived populations deviate frequently strongly from the base populations, but they are split into mostly different sets of positive and negative deviations.

Steps 3 through 5

For each of the 45 remaining ILMDd, I determined the set of "correlated" ILMDd among the other 44 remaining ILMDd (see above). In order to focus on composite traits, I discarded at this point 16 ILMDd which were not "correlated" with any other ILMDd. From the resulting list of 29 sets each assigned to one ILMDd, I further eliminated 9 sets which were subsets of

other sets and thus considered redundant. I thus ended up with the following list of 20 sets (the ILMDd notation "5-6-plus" means "increase of the ILMD between landmarks 5 and 6 in the derived population with respect to the base population"):

1-14-plus: 4-6-plus, 6-7-min

2-4-plus: 3-4-plus, 3-14-min, 5-6-min, 10-11-min

3-4-plus: 2-4-plus, 8-9-plus, 8-10-plus, 2-14-min, 3-14-min, 5-7-min, 5-8-min, 5-9-min

4-5-plus: 3-5-plus, 6-7-min, 12-14-min

4-6-plus: 1-14-plus, 2-5-plus, 3-6-plus, 6-7-min

5-6-plus: 5-7-min, 5-8-min, 12-14-min

8-9-plus: 3-4-plus, 8-10-plus, 8-12-plus, 11-12-plus, 5-9-min

8-10-plus: 3-4-plus, 8-9-plus, 2-14-min, 3-14-min, 5-9-min

8-12-plus: 8-9-plus, 9-10 plus, 11-12-plus, 6-9-min

9-10-plus: 8-12-plus, 4-9-min

11-12-plus: 6-7-plus, 8-9-plus, 8-12-plus, 2-14-min, 5-9-min

2-14-min: 3-4-plus, 8-10-plus, 11-12-plus, 3-13-min, 3-14-min, 5-9-min

3-14-min: 2-4-plus, 3-4-plus, 8-10-plus, 2-14-min, 5-7-min, 5-8-min, 5-9-min, 10-11-min

5-7-min: 3-4-plus, 5-6-plus, 3-14-min, 5-8-min, 5-9-min, 10-11-min

5-8-min: 3-4-plus, 5-6-plus, 3-14-min, 5-7-min, 5-9-min

5-9-min: 3-4-plus, 8-9-plus, 8-10-plus, 11-12-plus, 2-14-min, 3-14-min, 5-7-min, 5-8-min, 10-11-min

6-7-min: 1-14-plus, 4-5-plus, 4-6-plus, 12-14-min, 13-14-min

6-9-min: 8-12-plus, 9-11-plus

10-11-min: 2-4-plus, 3-14-min, 5-7-min, 5-9-min

12-14-min: 4-5-plus, 5-6-plus, 6-7-min, 13-14-min

Steps 6 through 8

A score for each set on the list was calculated for each base population and derived population from the raw ILMD data. The correlations between the scores across the populations (SUPPLEMENT 9) were transformed into distances and subsequently used to visualize the relations between the "leading ILMDs" on the list as a Neighbour-Joining topology (Fig.17).

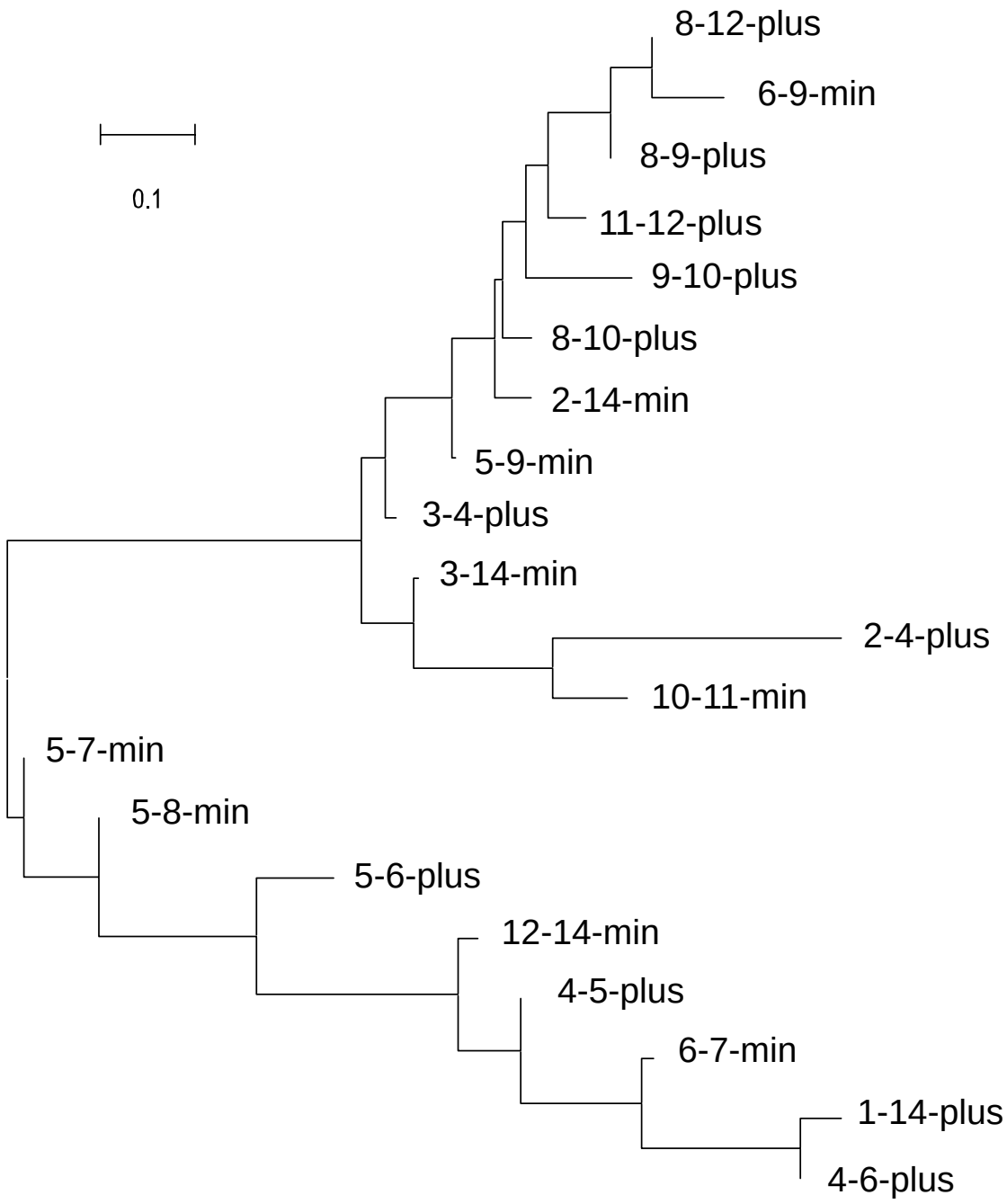


Fig.17. Neighbour-Joining tree based on the correlation matrix of the preliminary trait scores.

The topology shows that there is a distinctive correlation structure among the “preliminary traits” from the list. There are two large branches. The sets of “preliminary traits” occupying these branches are more or less unrelated. The “preliminary traits” within the large branches show a graded pattern of association.

Chapter eight – evaluation of the major trait hypothesis

In the previous chapter, I argued that the shape features of the derived populations can be used to identify structures of variability, which are hypothesized to correspond to "developmental traits". The limited number and partially conflicting directions of these traits impose constraints on the shape space, which manifest themselves as "parallelisms" in the "dysbalanced" shape space of the derived populations, where trait expression is expanded in comparison to wild mice for unknown reasons. An analysis of the dysbalanced shape space revealed two unrelated major traits, which in turn produce –in a hierarchical manner- a graded spectrum of correlated "subtraits", whose boundaries are not distinct. This hierarchy was hypothesized to correspond to a hierarchy of developmental processes.
From this hypothesis, we can derive the following expectations which can be tested in the context of the available shape data:

1) if the identified traits form the basis of the apparent "parallelisms" of the derived populations with respect to the base populations, the traits should explain large proportions of the differences between base and derived populations, it should be possible to obtain the parallelisms based on the traits alone, and these parallelisms should disappear after removal of variation in these traits via regression;
2) if the traits are significant producers of variation, they should explain an important amount of variation also within and between populations;
3) if the traits correspond to developmental processes, they should explain important parts of longitudinal ontogenetic variation.

These three expectations will be tested separately in the following sections.

i. Definition of the final traits

Before beginning with the analysis, it is necessary to redefine the traits in order to be able to deal with them in a nonredundant, practical fashion. The "preliminary" traits in Fig.17 are not directly suitable, because their definitions are overlapping and their number -20- is impractically large. Therefore, I subdivide the topology in Fig.17 somewhat arbitrarily in subtraits as shown in Fig.18.

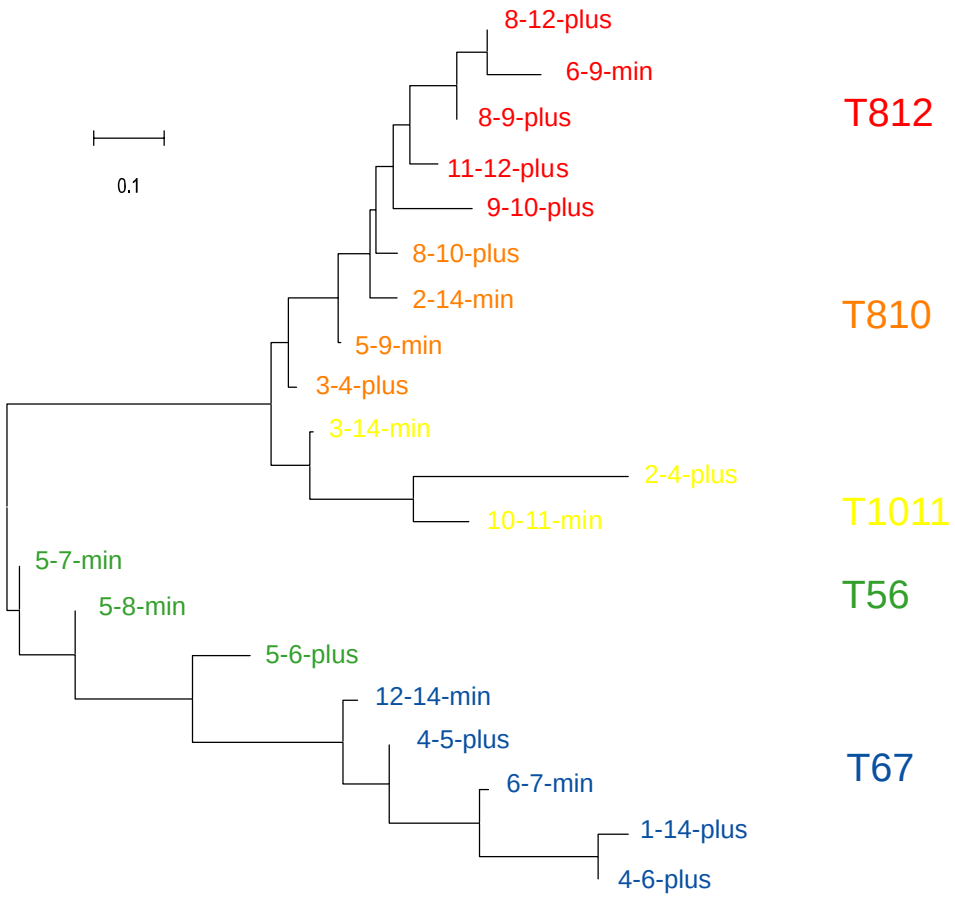


Fig.18.Definition of the 5 major traits which distinguish captive and inbred mice from wild mice. Note that the trait assignment is not supposed to reflect a hierarchical order, i.e. T810 and T56 are not "monophyletic", but reflect pairwise distance groups.

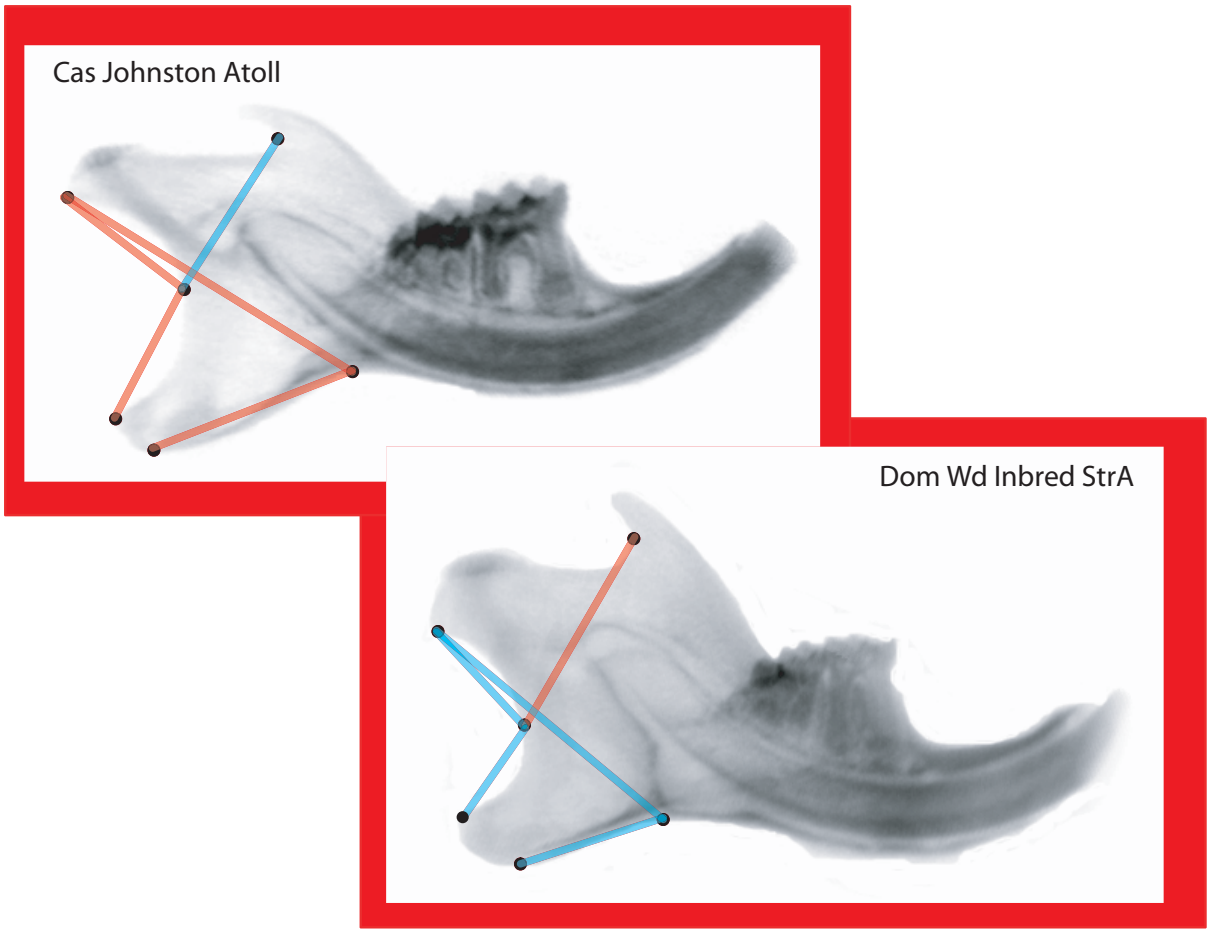
I thus obtained 5 final traits, which were used in the further analyses. It needs to be kept in mind, though, that the boundaries between the subtraits are imperfect solutions to the problem of subdividing a continuum of correlations. They therefore represent compromises rather than perfectly defined traits.

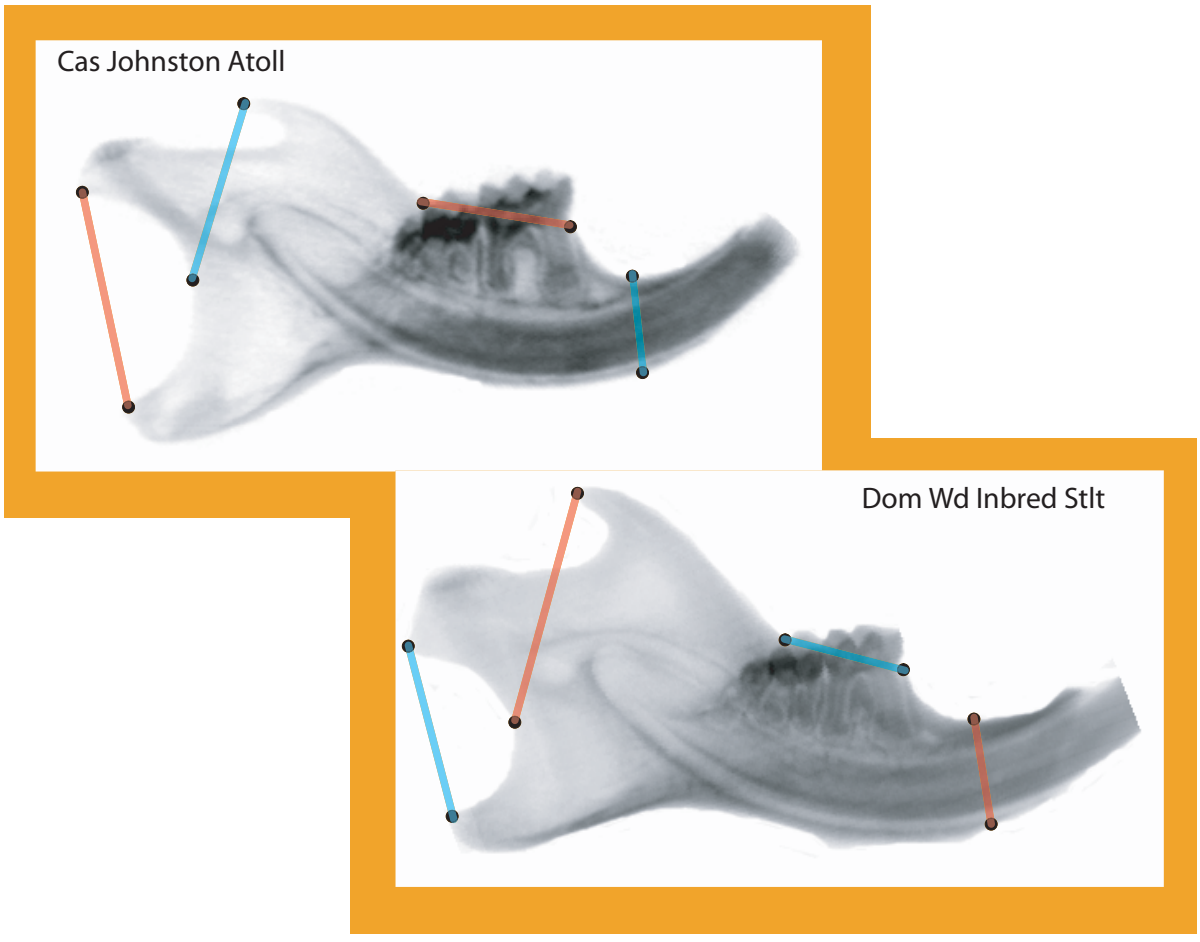
As it was done for the preliminary trait analyses, trait scores are defined as the sum of all positive ILMDs divided by the sum of all negative ILMDs. These scores can be calculated for population averages as well as for individual mice. Table 5 shows the correlations between the 5 trait scores for the population means of derived and base populations.

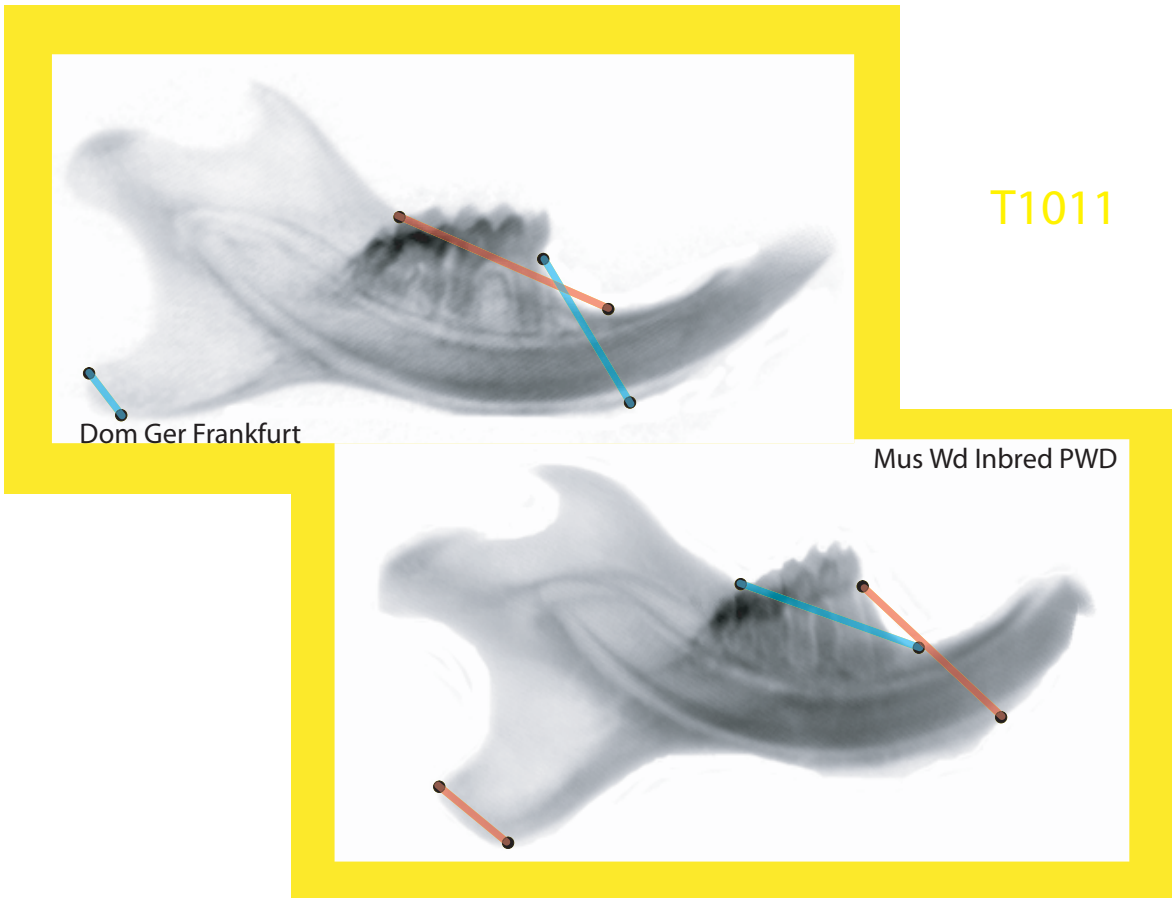
	T812	T810	T1011	T56	T67
T812		2.93E-06	0.17643	0.41618	0.15019
T810	0.79841		0.00422	0.58677	0.053622
T1011	0.28539	0.5625		0.22423	0.26822
T56	0.17398	-0.1168	-0.25762		0.049927
T67	-0.30292	-0.39871	-0.23537	0.4045	

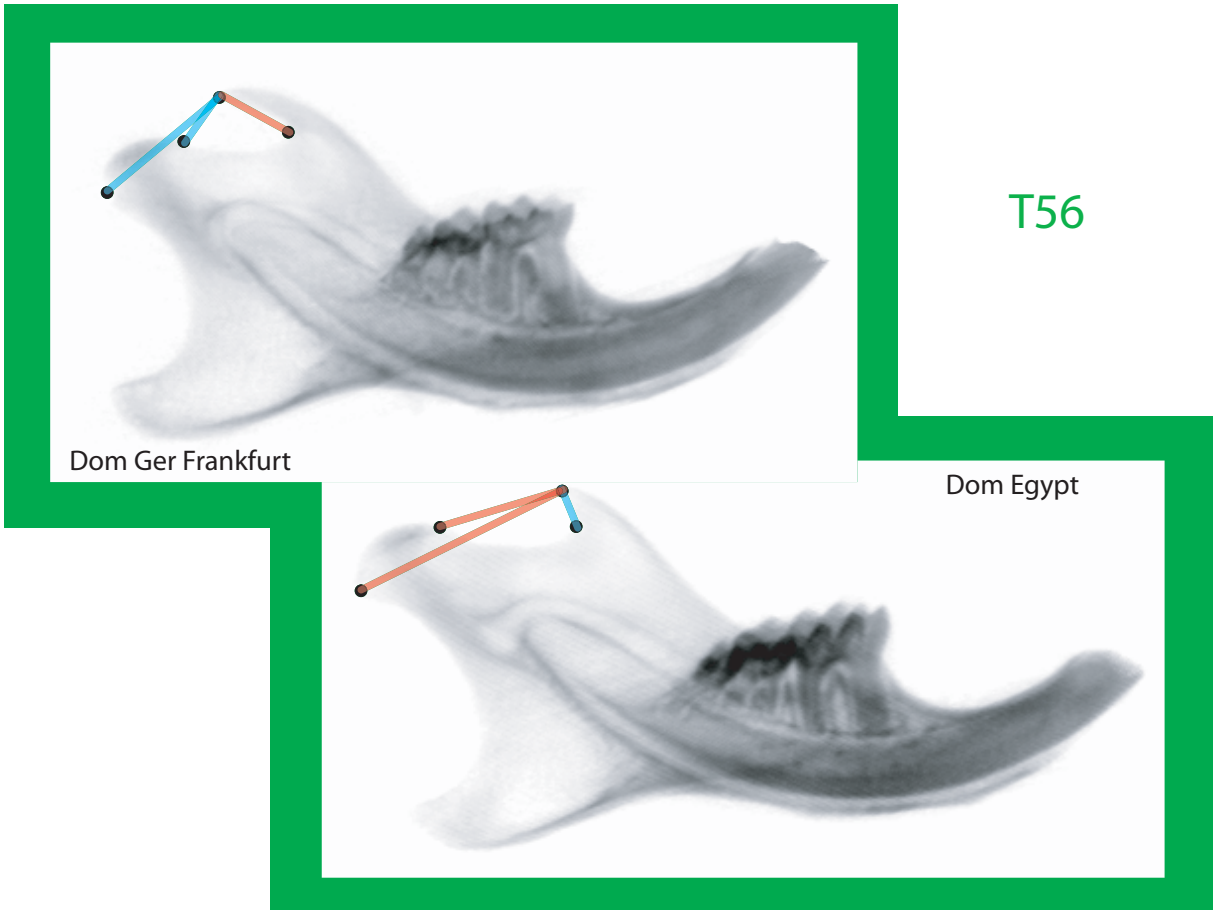
Table 5. Correlations between the trait scores of base and derived populations. The upper triangle gives the p values, the lower triangle the correlation coefficients.

From the table, it appears that the subdivision in 5 subtraits is a better solution than just keeping the two major branches, because the correlations between related subtraits are intermediate, indicating partial independence of the corresponding patterns of variation.
The 5 traits are shown in Fig.19.


Cas Johnston Atoll
Dom Wd Inbred StrA


Cas Johnston Atoll
Dom Wd Inbred Stlt


T1011
Dom Ger Frankfurt
Mus Wd Inbred PWD


T56
Dom Ger Frankfurt
Dom Egypt

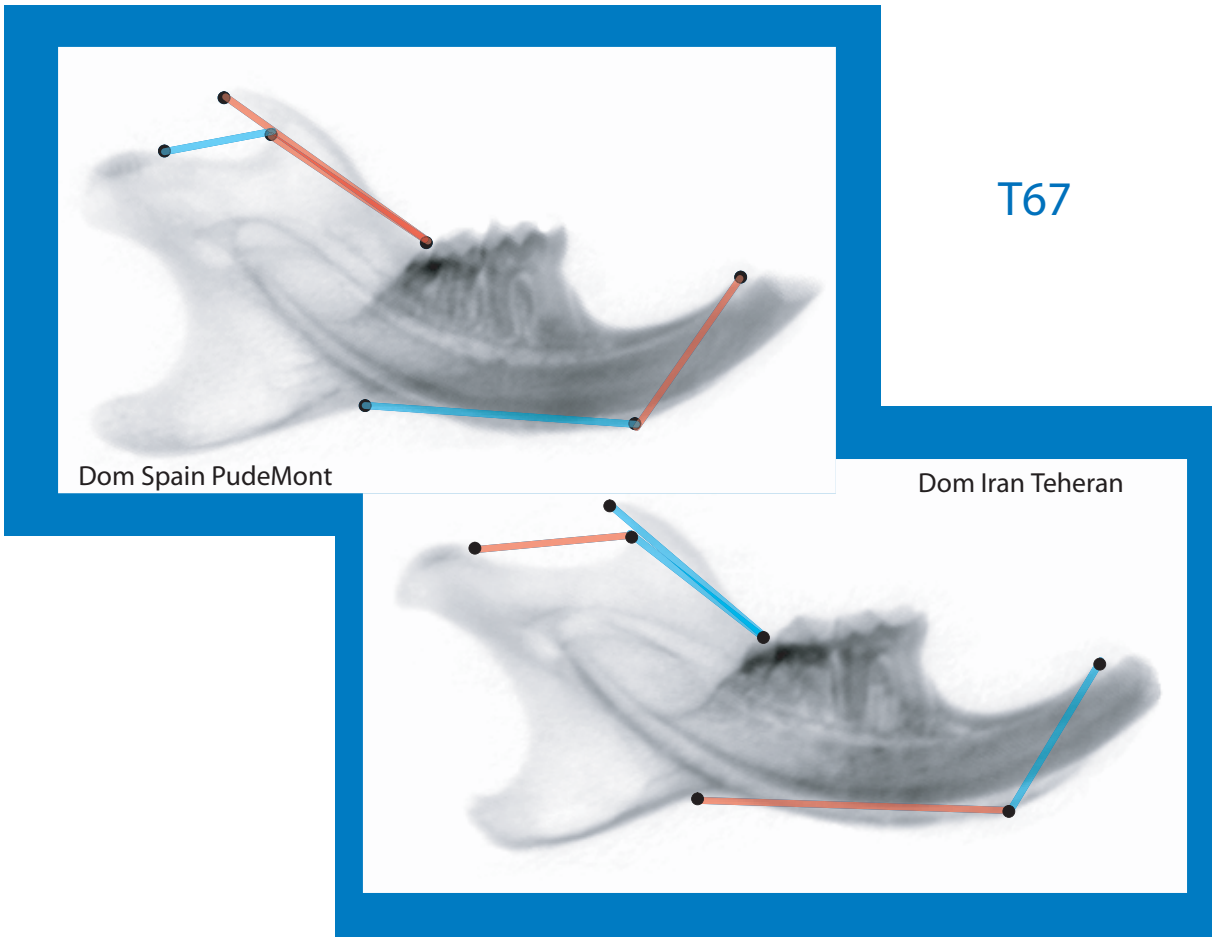


Fig.19. Visualization of the 5 traits. In each panel, the upper left panel shows a mandible with a high score for the trait, and the lower right panel shows a mandible with a low score. Red lines mark increased ILMDs and blue lines mark decreased ILMDs.

The traits may at least partly be interpreted as tradeoffs of material allocation. With an increase of T812, the angular process becomes larger and longer, while the height of the proximal condylar region decreases. This could mean that material from the condyle has been allocated to the angular process. Similarly, with a decrease in T810, the entire mandible seems to become higher, while the angular process is reduced in size. The decrease of ILMD 2-4 is apparently related to an increase in height of the region between the molar alveolus and the coronoid process, leading to a larger proportion of the posterior molar row being “hidden” by bone. From a visual comparison of T812 and T810, it becomes clear why they are to some degree correlated, because both of them correspond to the relative size of the angular process. As for T1011, it is not immediately clear why the height of the distal angular process is related to what appears to be mostly the height of the anterior molar row. There might be a functional correlation between these characters in the sense that a heightened molar row leads to a different insertion angle of the masseter for optimal biting force. T56 mostly represents the largely independent variation in length of the distal coronoid process. T67 consists of two

localized components: posteriorly, there seems to be a tradeoff between the length of the condyle and the size of the region between the coronoid process and molar row, including the proximal and distal coronoid process. Anteriorly, the ratio between the length of the proximal and the distal segment of the incisor ramus changes as a function of the changes in the posterior regions of the mandible. The relationships between the anterior and the posterior compartment of T67 may potentially be functional/epigenetic (this is speculation), whereas the changes within the compartments may represent growth tradeoffs. The correlation between T56 and T67 makes sense insofar as an increase in condylar length reduces by itself the relative length of the coronoid process in this region.

In summary, the mandible shape changes implied by the 5 traits appear to make sense in terms of growth tradeoffs and functional relationships, and their correlations appear to arise simply as a consequence of the mandibular geometry.

ii. Assessment of the importance of the traits for variation.

I used regression analysis to determine the importance of a trait as a feature of variation in a given context. To study variation within populations, I regressed either the Procrustes coordinates or PC scores of the Procrustes coordinates of a population onto the trait scores.

To analyse differences between populations or other groups, either CVA scores of a CVA including only the two groups in question or regression scores (subsequently called “shape difference scores”) from a regression onto a dummy variable encoding group membership were used as dependent variables in regressions on the trait scores. Both methods should give similar results.

The amounts of variance explained are then taken as an indication of how important a feature of variation is in the respective context. The amounts of variance explained from the joint regression of all 5 scores are to be taken with a grain of salt: they are meant to be a indicator of how important the 5 traits are altogether in a given context, yet the combined regression model by which they are calculated is not in itself meaningful here because the present purpose is descriptive, not causal. They are just the less problematic alternative to summing up the amounts of variance explained by the single trait scores.

Similarly, the proportion of cases in which the regression of Procrustes coordinates or shape difference scores on a trait score was significant in a given context is a measure of how important the trait is as a feature of variation in this context.

Finally, the amounts of variance explained for PC1 are an indication of how important a trait or traits are as determinants of the major direction of shape variation.

ii. The 5 traits explain parallelisms

The first expectation to be met if the 5 traits do fundamentally constitute the parallelisms between the derived populations with respect to the base populations is that the traits explain large proportions of the differences between derived and base populations. To quantify this, shape difference scores for each difference between a derived population and its base population were regressed onto the trait scores. Table 6 gives the amounts of variance explained.

	T812	T810	T1011	T56	T67	joint
BALB/cByJ	n.s.	n.s.	3.3	31.6	62.8	68
Cas Wd Inbred EiJ	50.6	46.3	8.3	42.8	42.8	77.7
Dom Lab Ger Cologne	43.1	56.2	16.7	*3.9	*9	69.5
Dom Lab France Gen3/4	44.9	65.1	30.7	n.s.	*15	81.1
Dom Lab Iran Ahvaz Gen1	2.5	9.4	32.2	n.s.	*22.2	45.5
Dom Kerg Guillou	53.7	35.6	n.s.	56.4	28.7	87.5
France x Iran F1	41.7	62.4	67.87	*6.2	*51.6	91.1
C57BL/6J x PWD F1	*2.3	34.4	13.8	7.1	9.3	54
Dom Kerg Cochons	n.s.	5.6	5.8	45.9	32.7	63.1
Mus Lab Khaz Gen1	4.1	*11.3	14.7	n.s.	41.7	52.7
Mus Wd Inbred PWD	7	16.6	50.5	41.3	9.9	83
Spr WD (Inbred) Spain Madrid	*20.5	67.2	*22.7	*28.7	8.3	76.3
Dom Wd Inbred Stlt	48.4	72.2	n.s.	30.2	n.s.	82.4
Dom Wd Inbred StrA	41.6	40.7	13.2	11.5	*7	59
Dom Wd Inbred StrB	12.8	*18.7	16.7	*6.8	*28.6	52
Dom Lab Iran Ahvaz Gen3	n.s.	40.4	46.5	*17.2	*40.3	78.4
Cas Wd (Inbred) Tai	24.1	48.7	31.4	n.s.	4.3	61.5
FVB/NJ	*3.4	11.1	7.1	52.7	77.7	89.2
C57BL/10J	n.s.	n.s.	n.s.	26.1	75.4	76.7
C57BL/6J x Cast/EiJ F1	n.s.	37.6	45.4	*32.3	*42.5	82.1
C57BL/6J	n.s.	*19.6	4.7	24.5	53	67.8

Table 6. Percent of variance of shape difference scores (see text) between base and derived populations explained by regressions (separately and jointly) onto the 5 trait scores.

The 5 traits do jointly explain on average more than 70% of the shape differences between base and derived population. This demonstrates that there are preferred directions of departure of the derived populations from the shape space of the base populations.

The second expectation is that it should be possible to reconstruct the "parallelism tree" in Fig.16 based on the trait scores alone. I tried this in the following way: after calculating the trait scores from the average ILMDs of the derived populations, I standardized each score by its mean and subsequently calculated the correlation matrix between the standardized score vectors of the populations (Supplement 10). This matrix was transformed into a distance

matrix and used to calculate an NJ-tree. The resulting tree is shown in Fig.20, together with histogram representations of the traits causing the pattern (see figure legend).

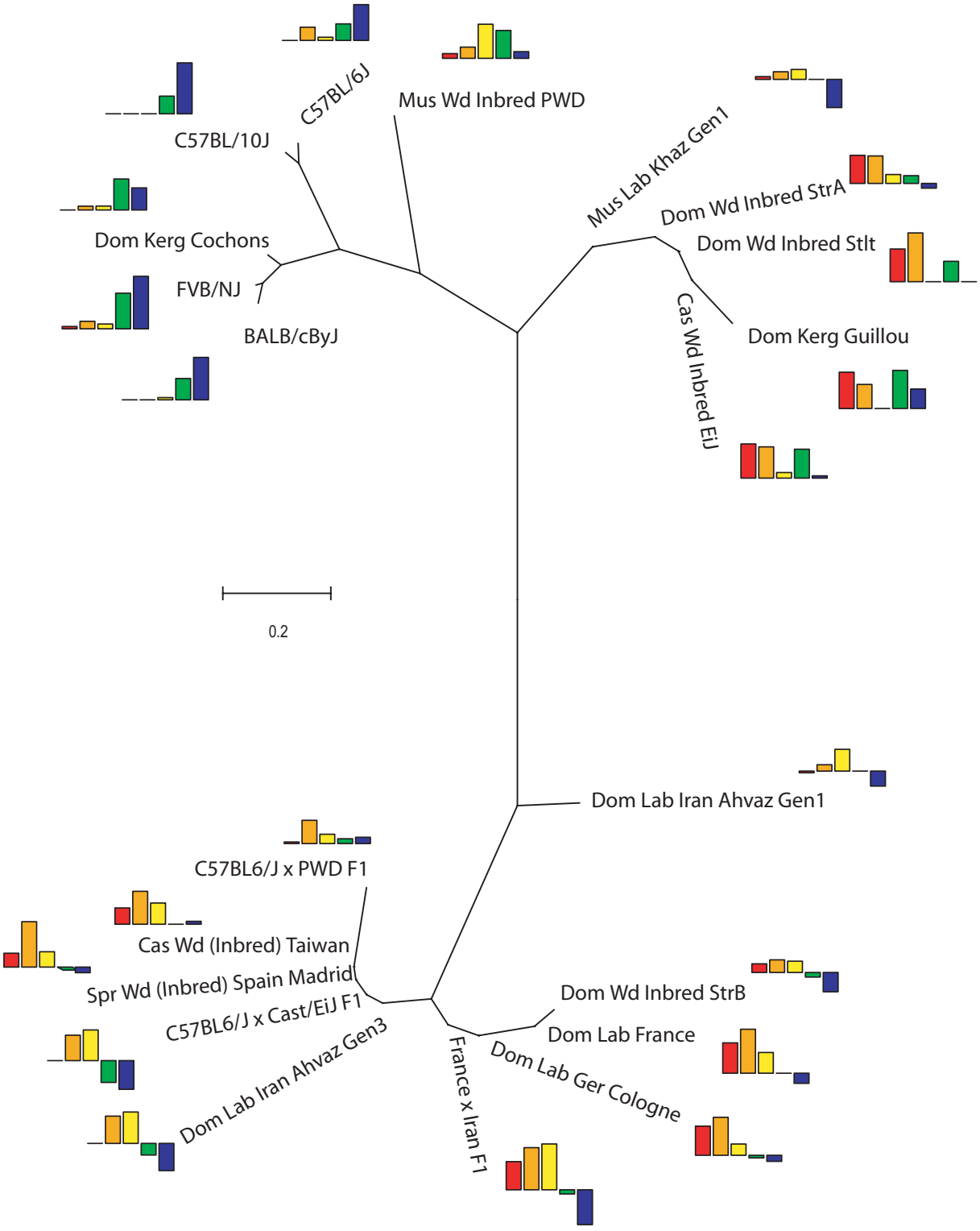


Fig.20. Neighbour-Joining topology explaining the relations between captive and wild mouse populations in the way they differ from the base populations by means of the 5 traits (see text for explanation). The histograms show the percent of variance in shape difference score explained by each trait (see Table 6.). An upward bar means a decrease of the trait score in the derived population respective to the base population, a downward bar means an increase.

The close similarity between this topology and the topology of the original "parallelism tree" shows that the 5 traits are sufficient to explain most features of the original topology. There are a limited number of discrepancies:

- the 5 traits do not explain the affinity of C57BL/6J x PWD F1 to its parental population C57BL/6J in the parallelism tree;
- the affinity between CAS WD INBRED EIJ and MUS WD INBRED PWD in the original tree is not explained by the 5 traits;
- the placement of DOM LAB IRAN AHVAZ GEN1 in the original tree remains unexplained by the 5 traits;
- the placement of MUS LAB KHAZ GEN1 and DOM WD INBRED STRB based on the 5 traits alone is different from the original topology.

These exceptions demonstrate that some minor features constituting shape affinities between derived populations are not captured by the 5 traits; this is expected because the procedure used to identify these traits is necessarily very rough. Most of the information contained in the original tree is recovered, however, using just the 5 traits.

An additional observation that can be made on the amounts of variance explained is that the deviations of the derived populations from wild mice are mostly in the same direction for each trait. For T810, T812 and T1011, the deviations are almost always negative, and if they are positive, then only very slightly. For T56 and T67, the deviations from wild mice tend to be in the negative direction in wild-derived outbred populations and in the negative direction in inbred strains (see discussion).

The second expectation is that removal of variation in the 5 traits by regression eliminates the "parallelisms" between derived populations. This can be tested using geometric morphometrics. I calculated the trait scores for each individual in the base and derived populations. The scores were imported as covariates in MorphoJ, and after multiple regression of the Procrustes coordinates on the scores, the CVAs which had first indicated the "parallelisms" (Fig.8 and Fig.11) were repeated using the regression residuals. Fig.21 shows that after removal of variation in the 5 traits, not only the parallelisms have disappeared, but most of the shape differences between base and derived populations are collapsed (note that in Fig.21, the separation between *M. musculus* and *M. macedonicus*/*M. spretus* along CV1 is mostly conserved by the regression, for the wild as well as for the captive mice).

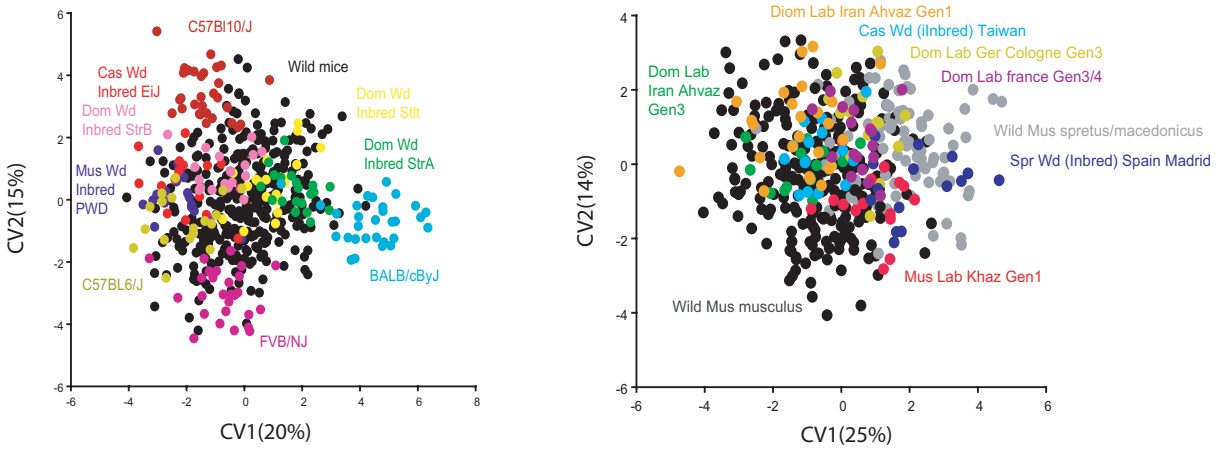


Fig.21. Scatterplots of the first two CVs of the wild mice of the reference dataset together with the inbred strains (A, compare with Fig.R.4.a.) and wild-derived outbred populations (B, compare with Fig.R.5.a.). The CVAs were calculated after regression of the Procrustes coordinates on the 5 trait scores jointly.

This shows that the 5 traits are not only responsible for the "parallelisms", they are responsible for most of the very shape distances between base and derived populations. The processes causing these distances in the first place are identical with the causes of "parallelisms".

iii. The 5 traits explain shape variance within and between populations and species.

If the 5 traits do structure variance in a similar way between derived populations and within wild mice, their relationships, i.e. their correlation matrix, should be similar to that in the derived populations. The same applies for the 20 preliminary traits on which the 5 traits are based. In order to see whether this is indeed the case, I calculated the trait scores for wild-caught *Mus musculus* and made a NJ-tree based on their correlation matrix (SUPPLEMENT 11) analogous to Fig.18 (Fig.22). It turns out that with the exception to the preliminary traits composing T810 and T1011, which are arranged in a slightly different way in wild mice, the relationships between the preliminary traits are identical between derived populations and indivdiual wild mice.

In addition, I calculated the correlation between both among-trait correlation matrices for the 5 traits: the one calculated from the scores of the population means of the derived populations, and the one calculated from the scores of individual wild mice. The correlation between both correlation matrices was 0.767. The relationships between the 5 traits are thus similar in both datasets. In the wild mice, however, the absolute values of the correlations are generally much lower than in the derived populations (Fig.22). This might indicate the presence of additional factors influencing shape and disturbing the signal obtained from the 5 traits.

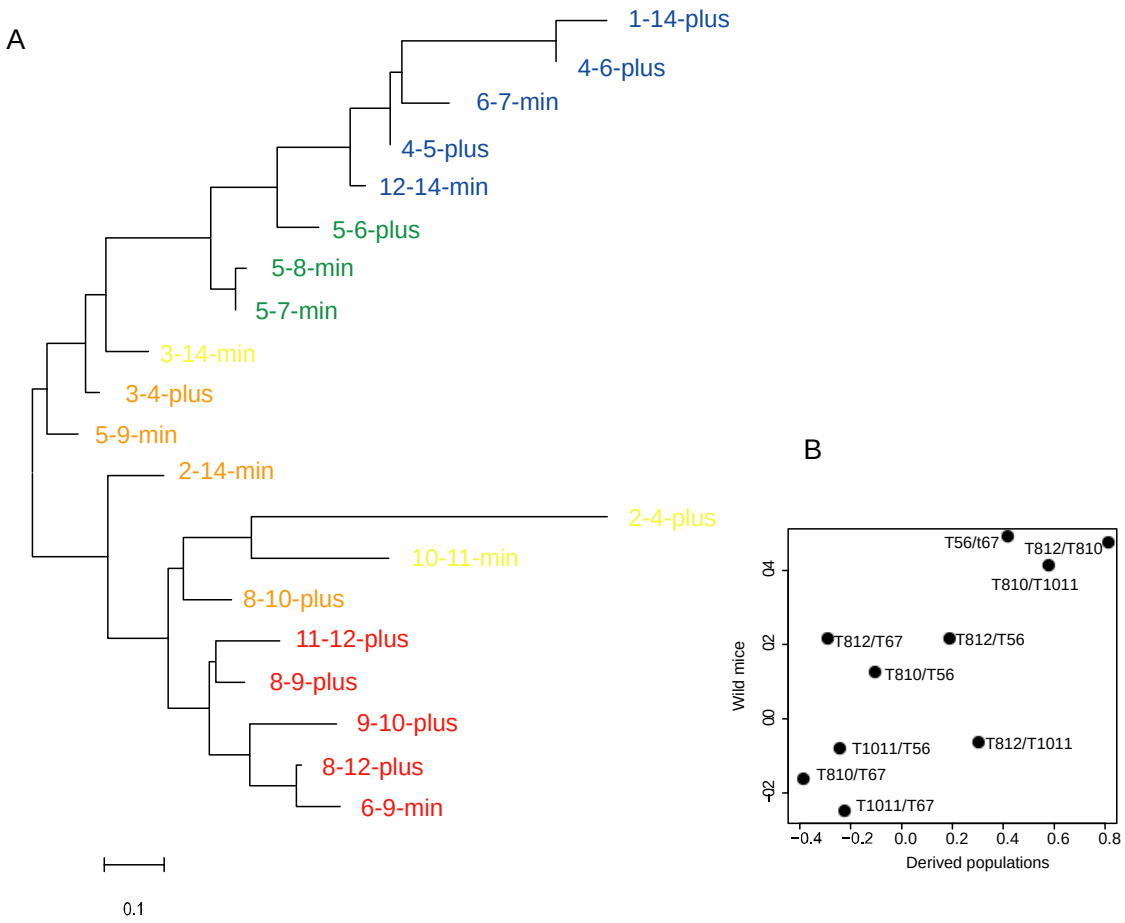


Fig.22. A, Neighbour-Joining tree based on the preliminary trait scores calculated for wild-caught *Mus musculus*. Compare to Fig.18. B, scatterplot of the correlation coefficients between the 5 trait scores for derived populations x axis) vs. wild mice (y axis).

To test whether the 5 traits explain significant amounts of shape variance within populations, I regressed the Procrustes coordinates on the trait scores. This was done separately for each population of wild mice, outbred captive and inbred mice. Because biologically relevant shape variation is expected to accumulate in the first few principal components, I determined additionally the proportion of variance explained for each shape PC1 in these populations.

The amounts of variance explained in these regressions are shown in SUPPLEMENT 12. I then determined the importance of the 5 traits for shape differences between populations and species. This was done using CVA scores or shape difference scores obtained as described above. These analyses were done for intraspecific differences between populations and for differences between species. The amounts of variance explained in these regressions are shown in SUPPLEMENT 13.

In order to abbreviate the discussion of results and facilitate understanding, I summarized the results for the percent of variance explained within and between populations and species in

table 7 which shows the proportion of cases in which the 5 traits explained significant amounts of variance and the average amounts of variance explained in these cases.

		T812	T810	T1011	T56	T67	Joint
Within populations							
Wild *M. musculus*	% sign. total	66.7	77.8	44.4	100.0	100.0	100.0
	aver. total	10.8	9.2	10.3	21.5	21.1	57.6
	% sign. PC1	11.1	11.1	22.2	77.8	88.9	100.0
	aver. PC1	14.8	22.9	24.6	67.3	58.2	84.6
Wild *M. macedonicus*	% sign. total	50.0	100.0	0.0	100.0	50.0	100.0
	aver. total	14.1	17.6	0.0	18.8	16.6	57.3
	% sign. PC1	0.0	100.0	0.0	50.0	50.0	100.0
	aver. PC1	0.0	49.7	0.0	63.3	36.6	83.9
Wild *M. spretus*	% sign. total	100.0	100.0	100.0	100.0	100.0	100.0
	aver. total	7.5	10.0	8.0	15.3	11.9	45.8
	% sign. PC1	0.0	0.0	0.0	100.0	50.0	100.0
	aver. PC1	0.0	0.0	0.0	31.0	27.7	52.9
Outbred lab *M. musculus*	% sign. total	50	16.7	33.3	100	83.3	100
	aver. total	16.3	9.3	23.9	24.5	25.0	62.8
	% sign. PC1	33.3	0	33.3	66.7	66.7	100
	aver. PC1	39.6	0.0	50.9	57.8	66.1	79.2
Inbred lab and Kerguelen	% sign. total	45.5	81.8	45.5	90.9	90.9	100.0
	aver. total	13.3	10.5	9.6	16.4	18.4	52.1
	% sign. PC1	27.3	18.2	18.2	45.5	54.5	100.0
	aver. PC1	37.5	19.9	27.1	58.5	53.2	65.0
Between populations							
Wild *Mus*	% sign.	27.8	41.7	25.0	16.7	44.4	100.0
	aver. CV1	16.3	15.7	20.9	13.6	20.4	33.2
Wild-lab/Kerguelen	% sign.	71.4	90.5	85.7	81.0	95.2	100.0
	aver. diff. score	26.7	36.8	24.0	27.4	33.1	71.4
Between species	% sign.	100.0	33.3	100.0	0.0	100.0	100.0
	aver. diff. score	8.8	15.8	12.0	0.0	8.1	26.5

Table 7. Within populations: average percent of variance explained of Procrustes coordinates (black numbers) and PC1 of the Procrustes coordinates (magenta numbers). The red and blue numbers give the proportion of cases in which the regressions were significant; the averages given are calculated only from these. Between populations: black numbers: average percentage of variance explained of shape difference scores (see text) or CVA scores; red numbers: proportion of cases in which the regressions were significant; the averages shown were calculated only for the significant cases.

The following points can be derived from this summary table:

1) The total and relative importance of the 5 traits in within-population variation differs between species. In *M. musculus* and *M. macedonicus*, the 5 traits do jointly explain similarly large amounts of variation (57% total and 84% of PC1). Yet, whereas in *M. musculus*, the most important traits are T56 and T67, the latter is of lesser importance in *M. macedonicus*, in which much more variation is explained by T1011. In *M. spretus*, the 5 traits are less important: only 45% of variance are explained in total, and

only 52% of PC1. All traits explain significant, but low amounts of variation. T56 and T67 are most important, but much less so than in *M. musculus*.

2) In *M. musculus*, the relative contributions of the traits to within-population variation seem to have changed with respect to the wild situation. It has to be remembered, however, that these conclusion are based on only six populations of outbred mice. In these, compared to wild *Mus musculus*, T812 and T1011 are more frequently and more strongly associated with PC1, while variation in T810 plays a much smaller role in lab outbred mice than in wild mice. The importance of T56 and in T67, as well as of the joint traits, is similar in both categories.
3) The 5 traits explain similar large amounts of total shape variation in inbred mice as in wild mice. Since there is not much genetic variation in inbred strains, this means that this variation must be generated by the developmental system randomly (noise) or in response to unknown epigenetic factors. The relative importance of T56 and T67 is smaller than in wild mice.
4) The 5 traits are of lesser importance in explaining shape differences between populations. No single trait is in more than half of the cases significantly associated with an interpopulation difference and on average, only about a third of the variance of the interpopulation difference shape scores is explained by the 5 traits. This might simply mean that populations have "a harder time" to differ in traits in which they are by themselves highly variable.
5) As already mentioned above, the difference between any captive or inbred lab population and its base population is to the most part (about 70%) percent explained by the 5 traits, which is the reason that they were defined ("found") based on this context.
6) Between species, even less of the differences are explained by the 5 traits. Especially T56 is totally irrelevant for interspecific differences. The explanation might again be that it is hard to differentiate in highly variable traits.

iv. Outcrosses

If the 5 traits are important as major directions of developmental and epistatic variance, as suggested by the "major trait hypothesis", then not only differences betweeen inbred strains should be explained by them, but also the "epistatic deviations" reported in chapter 6. This can be demonstrated in two ways.

First, it can be asked whether the "epistatic deviations", i.e. the vector of differences of the F1 from the interparental mean, correlates with the ILMDd vectors of differences between base and derived populations. If the developmental varinaces expressed in the two contexts are similar, then such correlations should exist. For each outcross, I calculated the vector of "epistatic deviations" by subtracting the ILMD vector of the F1 from the midparental ILMD vector, which is the mean of the two parental ILMD vectors. I then looked for correlations between the "epistatic deviation" vectors and the ILMDd vectors as explained above. The results are shown in table 8.

	ED: C57BL/6J x Cast/EiJ F1	***ED: C57BL/6J x PWD F1***	***ED: France x Iran F1***
C57BL/6J	0.501	0.061	0.348
C57BL/6J x PWD F1	0.049	0.334	0.031
C57BL/6J x Cast/EiJ F1	0.748	0.598	0.698
Cas Wd Inbred EiJ	0.384	0.528	0.182
France x Iran F1	0.382	0.216	0.784
Mus Wd Inbred PWD	0.458	0.726	0.023
Dom Lab Iran Ahvaz Gen1	0.038	0.08	0.063
Dom Lab France Gen3/4	0.117	0.037	0.536
BALB/cByJ	0.317	0.177	0.267
FVB/NJ	0.572	0.402	0.297
C57BL/10J	0.428	0.042	0.53
Dom Kerg Guillou	0.529	0.418	0.088
Dom Kerg Cochons	0.396	0.293	0.359
Dom Wd Inbred Stlt	0.456	0.26	0.326
Dom Wd Inbred StrA	0.369	0.307	0.219
Dom Wd Inbred StrB	0.09	0.117	0.39
Dom Lab Iran Ahvaz Gen3	0.565	0.461	0.526
Dom Lab Ger Cologne	0.264	0.142	0.639
Mus Lab Khaz Gen1	0.073	0.106	0.079
Spr WD (Inbred) Spain Madrid	0.58	0.4	0.57
Cas Wd (Inbred) Tai	0.459	0.296	0.705
ED: C57BL/6J x Cast/EiJ F1			
ED: C57BL/6J x PWD F1	0.754		
ED: France x Iran F1	0.609	0.331	

Table 8. Correlations between ILMD "epistatic deviation" vectors ("ED") and ILMDd vectors describing the differences between base and derived populations (see text for further explanations). The colours indicate ranges of p values: red: nonsignificant; yellow: significant at 0.05; green: significant at 0.001.

Most of the correlations in table 8 are significant. I conclude that the deviations of F1 from the interparental means are in similar directions as the differences between base and derived populations, which have been shown to be heavily dependent on the 5 traits, and that the same is thus likely to be the case for the "epistatic deviations".

The second way to assess how the 5 traits influence the differences between parentals and F1 in outcrosses is to regress the Procrustes coordinates on the trait scores ant to see how this affects the distance relationships in a CVA. I did this for the C57BL/6J x Cast/EiJ and for the C57BL/6J x PWD outcrosses. Recall from chapter 6 that in these outcrosses the parentals and the F1 form a more ore less equilateral triangle in a CVA plot (Fig.14, leftmost panels; in a CVA including three groups, the scatterplot of CV1 and CV2 summarizes the complete information). This means that the distances between the three groups are more or less equal. If the "epistatic deviations" result from superimposed differential contributions of the 5 traits to the differences between each parental and the F1, then regressing the Procrustes coordinates on the trait scores and calculating the CVA from the regression residuals should affect the shape of this triangle in various ways. Regression on all traits jointly should eliminate the "epistatic deviations".

Fig.23 shows the results of these analyses for the two outcrosses.

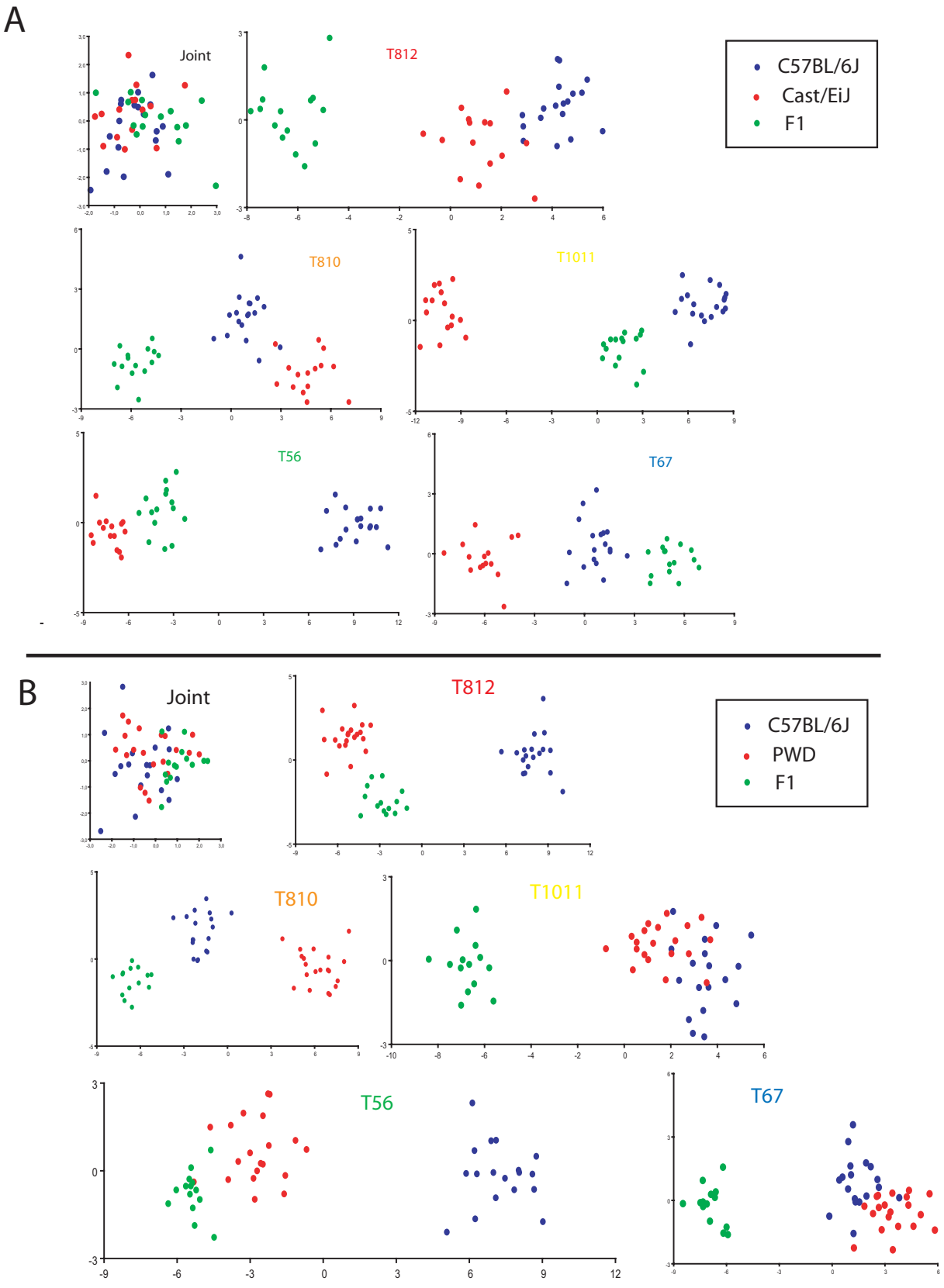


Fig.23. CVA scatterplots after regression of the Procrustes coordinates on the 5 trait scores (jointly, "Joint", or individually, as indicated for each plot). A: C57BL/6J x Cast/EiJ; B, C57BL/6J x PWD.

In both Outcrosses, regression on all 5 trait scores jointly removes all differences between the parentals as well as between parentals and offspring. In the regressions on individual traits,

some traits do specifically reduce the difference between the parentals (T812 and T810 in the the C57BL/6J x Cast/EiJ outcross, and T1011 and T67 in the C57BL/6J x PWD outcross). Other traits, however, specifically reduce the difference between one parental and the F1 (T1011, T56 and T67 in the C57BL/6J x Cast/EiJ outcross, and T812, T810 and T56 in the C57BL/6J x PWD outcross).I infer that the differences between parentals and F1 and thus the "epistatic deviations" are determined in a specific and differential fashion by variation in the 5 traits.

v. The 5 traits explain longitudinal ontogenetic shape change.

In order to evaluate this expectation, I made use of a previously generated dataset of two pseudolongitudinal ontogenetic series involving C57Bl/6J and MUS WD INBRED PWD mice. 15 mice of each strain were examined at 3 age stages: 2 weeks, 4, weeks, and 8 weeks age. Within each strain, I regressed the CVA scores of the 2 stage-wise comparisons (2 to 4 weeks, pre-weaning, and 4 to 8 weeks, post-weaning) on the trait scores of the individuals to see whether the traits were important factors of ontogenetic change. Furthermore, at each age stage, I regressed the CVA scores of the between-line comparisons on the trait scores, to see whether the traits contributions to between-strain disparity changed throughout ontogeny. The amounts of variance explained are shown in table 9, a graphical representation of the results in Fig.24.

	Joint	T812	T810	T1011	T56	T67
B6J 2 to 4 weeks	80.4	n.s.	15.8	75.5	n.s.	n.s.
B6J 4 to 8 weeks	31.3	n.s.	n.s.	n.s.	n.s.	18.5
PWD 2 to 4 weeks	88.1	50.9	n.s.	71.2	n.s.	18.4
PWD 4 to 8 weeks	71.0	n.s.	43.0	67.1	n.s.	54.3
B6J PWD 2 weeks	56.9	n.s.	25.8	n.s.	n.s.	n.s.
B6J PWD 4 weeks	82.0	13.2	58.8	n.s.	29.1	20.9
B6J PWD 8 weeks	88.7	n.s.	14.5	55.9	35.7	48.9

Table 9. Percent of variance explained by regression onto the 5 trait scores of CVA scores between ontogenetic stages within strains and between strains at the ontogenetic stages. n.s. = non significant. See text for further explanations.

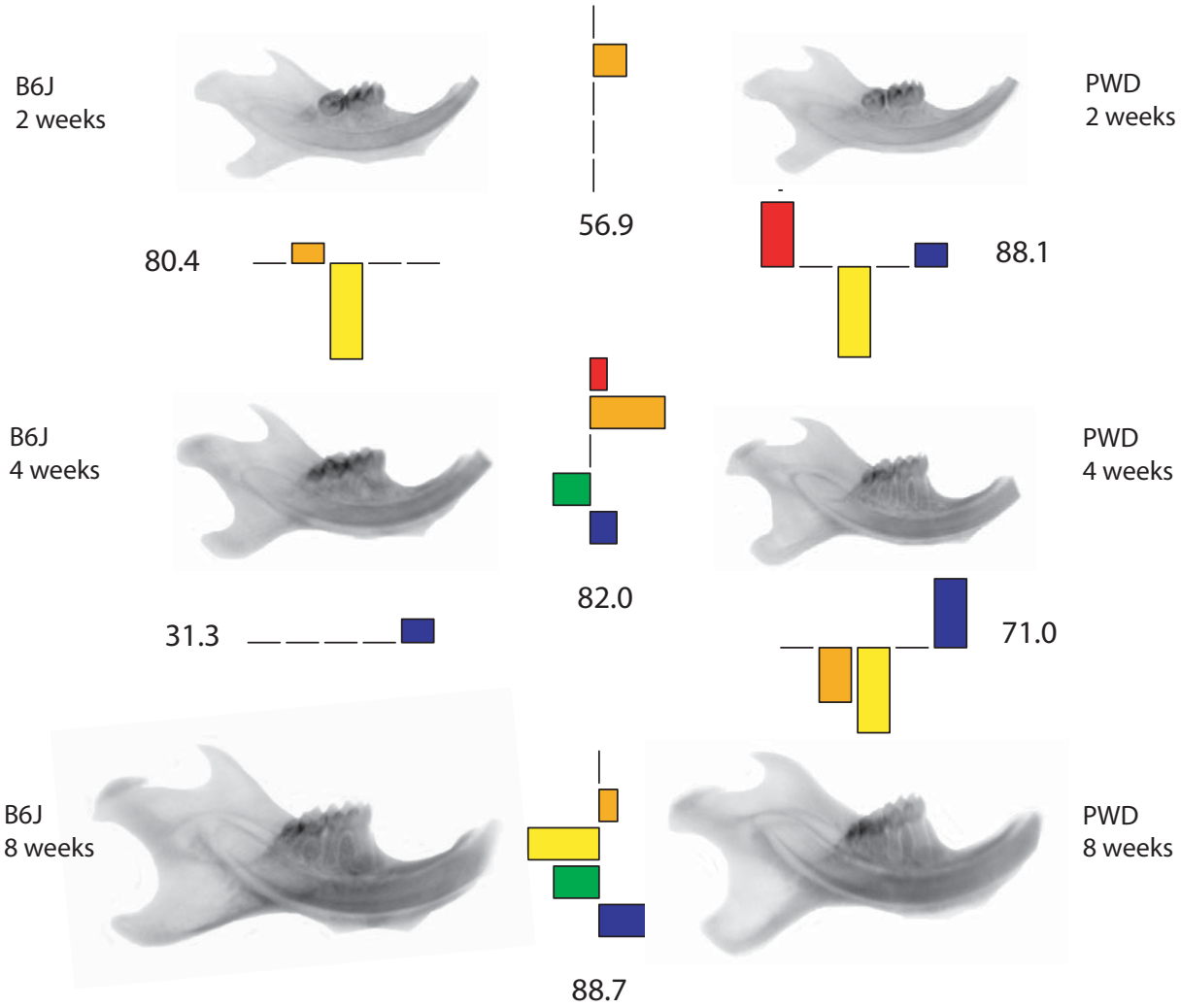


Fig.24. Qualitative visualization (see table 8.5 for absolute values) of the role of the 5 traits in the comparative ontogeny of C57BL6/J and MUS WD INBRED PWD between 2 and 8 weeks age. The C57BL/6J stages are sorted in the leftmost column, the MUS WD INBRED PWD stages in the rightmost column. within columns, longitudinal ontogenetic changes in the trait scores are shown: an upward bar means an increase in the trait score with age; a downward bar means a decrease. Between the strains, the vertical histograms show the shape difference between the strains at each stage. The bars point towards the strain with the higher score. The numbers indicate the percentage of variance explained jointly by the 5 traits.

In C57BL6/J, the 5 traits explain 80.4% of the pre-weaning shape change between week 2 and 4. This effect is almost entirely due to a decrease in the score of T1011. Only 31.3% of post-weaning shape change are explained by the 5 traits, mostly by an increase in the score of T67. The contribution of the 5 traits to MUS WD INBRED PWD postnatal ontogeny is larger and more complex. Here, the score of T1011 decreases between week 2 and 4 similarly as in C57BL/6J, and the scores of T812 and T67 increase. After weaning, the T1011 score increases such that its decrease before weaning is reversed, and there is a decrease in T810 and a further increase in T67, such that the 5 traits explain 71% of the post-weaning shape change in MUS WD INBRED PWD.

The development of the shape differences between C57BL/6J and MUS WD INBRED PWD reflects the different events in the ontogeny of both strains. At week 2, the 5 traits do already

explain 56.9% of the CVA score variance between both strains, although a clear effect is visible only for T810, where MUS WD INBRED PWD have on average a higher score. At week 4, the 5 traits do explain 82% of the CVA score variance. MUS WD INBRED PWD do now have higher values of T812, in accordance with the pre-weaning increase in this trait. The same pattern applies for T67. There is no difference in T1011 at this stage, probably because the values have decreased in both strains to the same amount before weaning. The difference in T810 between both strains has increased, and a difference in T56 with higher values in C57BL/6J has appeared. The latter two findings cannot be explained from the differences between MUS WD INBRED PWD and C57Bl/6J in 2 to 4 weeks pre-weaning development. Their appearance may be due to scaling effects: at 2 weeks, the mandibles are very small, increasing the effect of measurement error, such that shape differences may be difficult to detect (this is especially true for the very short ILMD 5-6). Additionally, imprecision of the CVA because of small sample sizes or undetermined growth processes might be responsible. At week 8, the varaince explained between boths strains by the 5 traits has slightly increased to 88.7%. The difference in T812 between both strains has disappeared, maybe because of other growth processes. The difference in T810 has decreased, probably due to the compensating decrease of this trait in MUS WD INBRED PWD between 4 and 8 weeks. The newly arisen difference in T1011with higher scores for C57BL/6J corresponds to the foregoing decrease in this score in MUS WD INBRED PWD. The difference in T56 has remained constant. The difference in T67 with higher scores in MUS WD INBRED PWD has increased, as MUS WD INBRED PWD has experienced a more dramatic increase in this score between week 4 and week 8 than C57BL6/J.

A large proportion of the shape difference between adult mouse strains can thus be seen as the end-product of complex differential choreographies of growth processes involving decreases and increases in the 5 traits.

Discussion of chapers 7 and 8

1. Identification of the 5 traits in the derived populations

Because of the methodological gap in geometric morphometrics in comparing and dissecting directions of shape change, I had to resort to using length measurements of distances between landmarks on the mandible outline (interlandmark distances – ILMDs). The full set of ILMDs which can be taken from 14 landmarks is a vector with 91 places. There is necessarily some redundancy in such a representation of mandible shape, but this does not pose a serious problem for the analysis. In any case, the ILMD vectors offer a simple and elegant way of comparing shape changes: just calculate the difference between a pair of shapes by subtracting the ILD vectors from each other, scale each place in the resulting difference vector by the average length of the corresponding ILMD to avoid confusion of shape and size, do the same thing for the second pair of shapes whose shape difference you want to compare with the first pair, and calculate the correlation between the two scaled ILMD difference vectors (ILMDd vectors). The correlation coefficient is then a measure of the relatedness of the shape differences or directions of divergence. While the geometry of the ILMDs in the mandible implies some redundancy in the ILMDd vectors, this is not a problem because it merely means that interesting shape changes are reflected in more than one element of the corresponding ILMDd vector. At the same time, the ILMDs offer the advantage that by simply sorting the ILMD vectors it is possible to spot the most important regions of shape change, whereas with geometric morphometrics, only complete landmark configurations can be considered due to the logic of the Procrustes superimposition.

In order to better understand the phenomenon of "parallelisms", I wished to quantify the relatedness of the directions of divergence of the derived populations from wild mice. This can be done using the methods described above and in chapter 7 and yields the "parallelism tree" of Fig.16. The original observation of "parallelisms" in the CVA plots is thus supported by a quantitative analysis. As can be seen in Fig.8, classical and wild-derived inbred strains diverge in sligthly different, yet related directions from wild mice. Most inbred strains sit on the end of relatively long branches compared to the wild-derived outbred populations, which form a rather tight cluster. Thus, whereas for the wild-derived outbred populations, speaking of a "parallelism" in the sense of very similar directions of divergence from wild mice makes sense, it does less so for the inbred strains, which do rather form a loosely connected cluster of more or less related directions of divergence.

The question is then how one can explain this “system of divergence”. Because the phenomenon as such is an entirely new observation, there is no preexisting theory to explain it. In this situation, I wish to bring up the idea that it results from a combinatorial superimposition of a small number of primary modes of shape changes brought about by the mechanistic substructure of the overall developmental process (e.g., more or less regionalized growth processes with specific directions which together determine the overall shape of the mandible). This relates remotely to heterochrony in the sense that there are supposedly a number of separate allometric growth processes whose expression can vary, for example, by differences in their specific rate and timing. If this hypothesis was true, all one would have to do would be to dissect the divergence vectors from a trait (ILMD) – based perspective, i.e. to find out which characters do regularly change in a coordinated manner. This amounts to a modularity perspective – not of modules as physically separate parts, but of modules as physically superimposable growth processes with specified directions. This differs from some applications of modularity within the morphological integration school, where modules are often construed as physically separate subregions of a morphological structure and their integration is investigated using statistical tools contrasting covariation between and within hypothetical modules (Klingenberg 2009).

As described in chapter 7 of the results part, conventional statistical methods were not useful for the required dissection of the divergence system. My manual procedure may be crude, but I consider it to be justified by its success in achieving the presumed goal: to identify a number of traits which can be considered separately, but taken together are sufficient to reconstruct and thus “explain” the “system of divergence”. This success becomes apparent in a comparison between Fig.16 and Fig.20. The topology of the “parallelism tree” is almost entirely recovered using just the correlations between the 5 trait scores across the derived populations.

Given that the “5 traits” are justified by the aforementioned result, it is necessary to first discuss their relationships and their nature, before applying them to better understand different contexts of shape variation. The final “5 traits” themselves are somewhat arbitrarily delineated within a topology of 20 preliminary traits (Fig.17 and 18). The preliminary traits are mutually overlapping groupings of ILMDs which were found to regularly appear together among all those ILMD which exhibit the largest differences between wild mice and derived populations. Each preliminary trait is defined by one ILMD and contains all other ILMDs which are regularly connected with this particular ILMD in the wild-derived differences. These preliminary ILMD groups are largely overlapping, which means that the unknown

underlying processes which produce this structure of variation are not confined to separate subregions of the mandible. The graded pattern of correlation among the preliminary traits as visible in the topology in Fig.17 is related to the degree of overlap between the ILMD sets, i.e. the more similar the regions affected by each preliminary trait are, the higher the correlation between the preliminary traits. This does, however, not imply a mosaic mode of modularity and development. Thus, traits may span several of the mandibular processus or link some of them with the molar row or parts of the incisor ramus (Fig.19). As detailed in chapter eight, the 5 traits can be interpreted as tissue allocation tradeoffs or more hypothetically as resulting from functional coupling, which may be developmentally controlled to yield a functioning mandibulofacial complex. This aspect does clearly need further detailed experimental studies of development. In any case, these results show that a physical mosaic perspective on developmental and genetic modularity is unrealistic in many respects and needs to be replaced by more biologically informed hypotheses. The 5 traits are furthermore rather unrelated to the "developmental units" of Atchley and Hall (1991), whose focus in explaining mandibular shape is on mosaic aspects of early development, during the formation of the initial mesenchymal condensations whose fusion gives rise to the mandible. As suggested by the results on the comparative developmental series of C57BL/6J and PWD in this thesis (see below), the 5 traits become manifest as patterns of variation mainly during postnatal ontogeny.

2. The 5 traits as components of shape variance

At this point, it becomes possible to integrate the ILMD results with geometric morphometrics by calculating trait scores and using them as covariates for regression analyses. The trait scores are useful because they are inherently scale-free and their purpose is to quantify for an individual mandible or the average shape of a population the "expression" of a given trait. Variation in each of the 5 traits can thus be independently described by these scores, which are based on nonoverlapping sets of ILMDs. There are two contexts in which regression on the trait scores can be used: either by using Procrustes coordinates as dependent variables, to determine how much variation within a population can be attributed to the traits, or by using a CVA or regression scores describing the shape difference between two groups as dependent variable, in order to determine how much of the shape difference between two groups relies on the traits. The results are quantified as "percent of variance explained". This is to be taken with a grain of salt: the "percent of variance explained" for different traits

cannot be added together, because the trait scores themselves may be correlated, and there is also a special caveat for the "joint" regression on the combined 5 trait scores at once: the underlying regression model assumes a causal relationship, where our purpose is just to quantify in a descriptive manner the importance of the 5 traits together in a given context. Therefore, the estimation of the importance of the 5 traits by the "percent of variance explained" is to be taken as a semiquantitative or nearly qualitative assessment.

With the 5 traits and a method to assess their importance in hand, I set out to test the following hypotheses (a-e):

a) The "parallelisms" observed in the derived populations are a consequence of the fact that the divergence is mainly caused by "overexpression" of just a few developmental traits. The "divergence system" or "parallelism topology" arises due to variation in a combinatorial "code" of "overexpressed" modular traits. The main result which confirms this hypothesis has already been mentioned, namely that it is possible to reconstruct the "parallelism tree" based on the 5 trait scores. A visualization of the amounts of shape difference explained by the 5 traits between each derived population and wild mice as a histogram of "amounts of variance explained" (Fig.20) gives an immediate impression of the idea of superimposition of modular "overexpressed" traits. The minor differences between the "parallelism tree" in Fig.16 and the "reconstructed parallelism tree" in Fig.20 are easily explained as a consequence of the relative roughness of the method used to find the traits and of the limitations of the dataset. The notion that the 5 traits are very powerful explanators of "parallelism" is reinforced by the result that they do jointly explain on average more than 70% of the shape difference between derived populations and wild mice, and that most of the divergence as well as the apparent "parallelisms" disappear from the CVA plots after regression of the Procrustes coordinates on the 5 traits (Fig.21).
b) From the hypothesis that the 5 traits correspond to growth processes in development, the expectation is derived that they should explain not only aspects of longitudinal shape change during ontogeny, but they should do so in a differential fashion between strains such as to explain the phenotypic differences between adult animals: "evolution is (genetic) change in development". I assessed this expectation by analyzing and comparing the postnatal shape change trajectories of two inbred strains, C57BL/6J and PWD. As described in detail in the last part of chapter eight, the expectation is fully

confirmed. Not only do the 5 traits explain substantial amounts of ontogenetic shape change before and after weaning, especially in PWD, but the differential ontogenetic shape changes in both strains do also add up nicely to explain the shape difference between adults (Fig.24), of which finally almost 90% is explained by the 5 traits. From a developmental point of view, it is interesting to note that the ontogenetic allometry is not a constant vector of shape change, but a flexible and dynamic process. Furthermore, the amount of shape difference between the strains explained by the 5 traits increases drastically between two and four weeks. Afterwards, it remains more or less constant, while the relative contribution of the traits is changing. These findings are not in conflict with the summary representation of the same data in

c) Based on the results for the first hypothesis, I hypothesize further that the 5 traits and their relationships as manifested in the "divergence system" of the derived populations do immediately reflect variational consequences of the fundamental properties of the developmental processes which define mandible shape in mice. This entails that the correlation among the 5 traits and among the 20 preliminary traits on which they are based are consistently found also in wild mice. It entails further that the 5 traits are important determinants of shape variation within wild mouse populations. The first prediction is shown to be true in Fig.22. The topology of the 20 preliminary trait scores in wild mice closely resembles the one derived from the "system of divergence" of the derived populations (compare with Fig.18). Consequently, the correlation structure among the 5 traits in wild mice likewise closely resembles the one previously found within the derived populations. The only slight difference is that the delineation of T812 and T1011 from each other in wild mice does not exactly follow the pattern in the derived populations. This may be due to a lack of resolution between these two closely related traits. Furthermore, the absolute values of the correlation coefficients are lower in wild mice (Fig.22). This does probably reflect the higher level of noise introduced by looking at individiual mice instead of population means, the lower level of total divergence and the noise introduced by the sampling limitations for wild-caught animals. The second expectation, that the 5 traits are important determinants of variation within wild populations, is borne out by the amounts of variance explained within wild populations and especially in PC1 of the Procrustes coordinates (Table 7). I will argue here mainly about *M. musculus*, because it is the focus of the study, and, with the exception of SPRE WD (INBRED) SPAIN MADRID, all derived populations are derived from this species. In *M. musculus*, the 5 traits explain on average 57.6% of the

total Procrustes variation within populations, and 84.6% of PC1. This means that the 5 traits are the most important directions of variation within wild populations. An important point can be added to this observation by considering the variation within inbred strains. I argued above that the properties of the developmental system are at the basis of this variation. Within inbred strains, which exhibit no genetic variation, the 5 traits do likewise explain large amounts of variance, albeit somewhat less than in wild-caught mice and in wild-derived outbred populations (Table 7). From this, two conclusions can be drawn: first, the developmental system produces structured variation in a manner partially independent from the input of genetic variation. In other words, the degree of expression of the traits is not completely under genetic control, but is subject to "uncontrollable variation" perhaps due todevelopmental instability. Second, the fact that in outbred populations, whether wild-caught or under laboratory condition, higher amounts of variation are explained by the 5 traits than in inbred strains, suggests the presence of genetic variation for their expression.

d) Having obtained the above results, I hoped to have found the key for explaining intra- and interspecific divergence in the sense of "lines of least resistance". I thus expected that the 5 traits play an important role for the explanation of differences between populations and species. This expectation was not met by the data. In fact, only 33% of intraspecific differences between populations and even less, 26% of interspecific shape differences can be attributed to the 5 traits (Table 7). This is in striking contrast to the average more than 70% of shape difference between wild mice and derived populations explained by the 5 traits. Although one third to one fourth of the shape differences explained are not unsubstantial, the major part of between-population and between-species differentiation is in directions unrelated to the 5 traits and thus to the major directions of variation within populations. A related finding is that the relative contributions of the 5 traits to the within-population variation seem to differ between species. Although *Mus spretus* and *Mus macedonicus* are only represented with 2 populations each in my dataset, the corresponding patterns in these species differ in a consistent way from *Mus musculus*. Thus, within populations of *Mus macedonicus* T810 explains much more variation than it ever does in *Mus musculus*, whereas T1011 seems to play no role in the former species, whereas it often does so in the latter. The importance of T67 may also be lesser in *M. macedonicus* than in *M. musculus*. In *Mus spretus*, less variance is explained by the 5 traits than in the former two species. The differences between the relative contributions of the traits are less pronounced, and

especially the contribution of T56 and T67 are reduced in comparison to *Mus macedonicus* and *Mus musculus*. These results indicate that the organization of the developmental system producing the within-population variation may in itself be subject to evolutionary change.

e) Above, I have argued that the divergence of the derived populations from wild mice is due to non-additive genetic interactions. A natural consequence from the finding that its directions can be mainly explained by differential "overexpression" of the 5 traits is the hypothesis that the non-additive genetic variance is mostly expressed via the 5 traits, or, to put it differently, that the 5 traits are the most important traits in the context of the non-additive genetic variance. This leads to the expectation that the "epistatic deviations" found in the F1 of outcrosses between inbred strains are also oriented alongside the 5 traits. This is already indicated by the high occurrence of correlation of the ILMD "epistatic deviation vectors" with the ILMDd vectors describing the shape differences between wild and derived populations (Table 8) and by the previous finding that outcrossing leads to a partial phenotypic reversal of these shape differences (Fig.14). The most striking demonstration comes, however, from CV analyses of the outcross triangles consisting of the two parental strains and the F1 after regression on the 5 traits. These analyses were done for the outcrosses of C57BL/6J with Cast/EiJ and PWD (Fig.23). Regression on all 5 traits jointly removes the differences between the parentals and the F1 almost completely in both outcrosses. This indicates already that all the epistatic variance is expressed via the 5 traits. The regressions on individual traits reveal the differential roles of the traits in explaining the differences between parentals and F1 within and between the two outcrosses (see details in Fig.23 and text in chapter eight). The observation is that specific traits distinguish between specific parentals and the F1 in each case (this is why these differences can be removed in a specific manner by regression on individual traits). The mechanistic interpretation is that the combination of the genomes of the parentals in the F1 activates in a nonadditive fashion a different combinatorial "trait code" than has been present in either parent. This new trait code leads to a phenotype which is somewhat more similar to wild mice – this is an epistatic complementation effect. A most promising perspective towards understanding the genetic architecture and development of mandible shape would thus be to explicitly map epistatic QTL for the 5 traits in such crosses between inbred strains.

An as yet unexplained finding is the bias in the directions of divergence of the derived population from wild mice for the 5 traits. For T810, T812 and T1011, the deviations are almost always negative, and if they are positive, then only very slightly. For T56 and T67, the deviations from wild mice tend to be in the positive direction in wild-derived outbred populations and in the negative direction in inbred strains (Fig.20). In the case of the wild-derived outbred populations, this pattern may be specifically associated with the hypothesized epigenetic mechanism of shape change from the wild mice (see above). In the case of the inbred strains, it may result from a developmental constraint.

Discussion part II – Synthetic view

The aim of this thesis was to describe the natural variation of shape of the house mouse mandible and to learn how it can be understood as resulting from the interaction of development and genetic architecture with microevolutionary processes in natural populations. I will now try to integrate in a broad perspective the insights from this work with knowledge and theory from the literature.

This work is unique in that it is an attempt to describe the overall shape space of an anatomical structure based on a broad sampling of populations from a species and its sister species (with the exception of Macholan 1996 whose interest was in taxonomy). Previous morphometric work has been restricted to more specific populations and comparisons. This broad sampling is crucially important because it allows generalization of the results to the entire species. For the first time we can now quantify the distribution of intraspecific and interspecific shape differences and shape variances. This was only possible after a careful assessment of the technical problems associated with combining heterogeneous material in a common analysis.

The knowledge thus gained about the natural distribution of shape differences between populations is the key to identifying interesting study populations and to asking the right questions about specific cases. For instance, we can ask why mouse populations from summer-dry regions are so similar in their mandible shapes and why the mice from the Kerguelen islands are so different from other house mice, and we thus have an opportunity to learn about the processes which influence mandible shape in general.

One of the most important key insights which became accessible through this approach was an assessment of the relative power of phenotypic plasticity in responce to different environments to change mandible shape. Phenotypic plasticity in the mouse mandible has traditionally been investigated using hard diet vs. soft diet experiments, and diet consistency has indeed been found to influence mandible growth and shape (e.g., Levrini et al. 2003, Maki et al. 2002, Yamada and Kimmel 1991, Renaud and Auffray 2009, Renaud et al. 2010). However, the magnitude of the shape change introduced by the difference in diet consistency in relation to the shape differences between natural populations has not been addressed in these studies. Through a comparison with the range of natural shape differences between populations it becomes possible to state that plasticity is not sufficient to explain the population level of phenotypic diversity in mouse mandible shape. The fact that a similar diversity is found between wild-derived outbred populations under identical laboratory

conditions and the additional evidence from comparisons of variance and ontogeny discussed above allow a firm establishment of the primary importance of genetic differences to explain phenotypic differences between natural populations.

This raises two questions: 1) why are the populations phenotypically distinct beyond possible effects of plasticity, and 2) what limits this diversity?

The obvious answers are: diversity is driven by positive selection or drift, and the shape space is limited by stabilizing selection. These answers are probably correct, but in their generality they hide the more subtle, but also more concrete problems raised by the structure of the natural shape space: if conspecific populations are different because they occupy different adaptive peaks with respect to mandible shape, why are they not more distinct from each other? In comparisons of conspecific populations, on average 15% of specimens are misassigned in discriminant function analyses (20% if one leaves out the highly distinct Kerguelen mice), and this takes into account the entire shape information. As can be seen in the CVA plots in Fig.3, even the most distinct populations overlap with other populations on the very axes describing the major directions of divergence. If the adaptive peaks are distinct, then why are the populations not more distinct? Why is the overall shape space so densely occupied? Why do house mice mostly not enter the shape space of the sister species, although the shape spaces are so close to each other (Fig.R.2.a.)? Are there such things as species niches/adaptive peaks with population subniches/peaks? Is stabilizing selection similar for all house mice, or is it diverse, or both? How important is neutral evolution? How precisely can mandible shape be genetically controlled, and, on the other hand, selected for (or against)? Is shape variation generated by mutation randomly distributed in shape space, or are there constraints? How is variation between populations related to variation within populations?

As can be seen from this list of questions, the secrets of phenotypic variation and of morphospace occupation within a species make up a largely unexplored field of empirical and theoretical problems which are intimately linked to questions of ecology on the one hand, and of genetic architecture and development on the other, with population and quantitative genetics right in between them. This entire field of problems is simply shielded from perception by sweeping references to adaptation, selection and sometimes drift.

In addition, conventional population genetic theory has so far been unable to solve the conundrum of how standing genetic variation for phenotypic traits is being maintained in the face of stabilizing selection (Barton and Keightley 2002). Studying actual phenotypic variation might just be the right place to go to better understand this problem.

A real eye opener in this context was the discovery that inbred mouse strains, which are genetically subsamples of wild populations (or entire species, in the case of the admixed "classical" laboratory strains) have mandible shapes which differ wildly from those found in nature. As described above, this does immediately invoke the hypothesis of a prevalence of non-additive genetic effects, especially epistasis, in the genetic determination of mandible shape. The fact that outcrossing betweeen inbred strains leads only to a partial phenotypic reversal to the "wildtype" shape underscores the specificity of the involved epistatic interactions in contrast to simple complementation of "recessive" alleles. This result has far-reaching consequences for understanding phenotypic variation in natural, genetically variable populations. If the phenotype of each individual is determined by a plethora of specific epistatic interactions, then a certain degree of phenotypic variation in the population is unavoidable as a consequence of mere recombination. Because the effects of any specific allele on the phenotype depend on the genetic background, stabilizing selection can only determine the phenotype with some degree of imprecision by removing alleles with strong deleterious effects on many different backgrounds. At the same time, the epistasis will shield many alleles from becoming eliminated by stabilizing selection, thus helping to maintain genetic variation.

Such variable interaction scenarios have been extensively discussed and modelled in the context of understanding the genetic architecture of the phenotype (Alvarez-Castro et al. 2009, Turelli and Barton 2006, Hermisson et al. 2003, Hansen et a. 2006, Wagner and Mezey 2000). Hill et al. (2008) suggested that the current empirical studies on variance components of complex traits provide little evidence for non-additive effects. On the other hand, studies that were specifically designed to reveal epistatic effects have confirmed epistasis between loci that were thought to act additively (Phillips 2008). Also, Moore and Williams (2009) have argued that additivity models do not appear to properly reflect the relationship between genotype and phenotype in genetic disease mapping studies in humans and suggest that epistasis and locus heterogeneity should be included to explain the genetic effects.

Before continuing on the theme of variation within populations, it is necessary to consider more specifically which kind of variation can be hypothesized to be produced predominantly in this way. It was first observed in CVA plots that the inbred mouse strains do not diverge in random directions from the wild shape space. As described above, this led to the discovery/islolation of 5 "traits", which are those directions of variation which are describing most of the divergence of the derived populations. These directions of variation were also found to determine the major part of shape variance within wild populations. This result

provides a link to the issues of canalization and developmental constraint, both subsumed in the idea of “lines of least resistance” (Schluter 1996).

Built on the (problematic, Pigliucci 2006) conceptual foundation of a stable G matrix, the idea of “lines of least resistance” predicts that the evolution of the phenotype will be along those directions in phenotypic space which correspond to the major eigenvectors of G, with G imposing a (developmentally rooted) constraint on the course of evolution. Epistasis is not being accounted for in this model (neither are fluctuations in G). Based further on the results of Cheverud (1988), implying similarity between G and the phenotypic covariance matrix P and thus the possiblity to substitute P for G, Renaud et al. (2006) apply the concept of lines of least resistance to fossil rodent teeth, with apparent success.

The divergence of the derived populations from wild mice in my data would be a perfect example of phenotypic change along lines of least resistance. The derived populations differ phenotypically from wild mice in a limited number of directions, and these correspond to the major directions of phenotypic variation within wild populations. However, these mostly bottlenecked/inbred populations probably cannot be considered as a standard model for the (long term) course of evolution. Furthermore, while the “lines of least resistance model” focuses on G and thus on additive genetic variance, the divergence of the derived populations is hypothesized here to result from epistasis.

My results on intra- and interspecific natural divergence are in direct conflict with the “lines of least resistance”. These directions of divergence disagree with the major directions of variance within populations. What could be the reason for this discrepancy?

A possible explanation relates to the idea that the specific phenotypic features of epistatic variance may be similar to specific phenotypic variance caused by developmental instability. The 5 traits, which were described based on the – epistatic- divergence of the derived populations from wild mice do not only explain major variation in natural populations, but also within *inbred* strains, where no genetic variation – epistatic or additive- is present. This variation must be due to developmental instability. The link between epistasis and developmental instability is further supported by the QTL results of Leamy et al. (2005), indicating that fluctuating asymmetry of tooth shape and size in mice – an indicator of developmental instability – has a predominantly epistatic genetic basis. The idea here is that developmental instability may be associated with sensibility to changes in epistatic interactions and may result in similar growth/shape changes.

If this would be the case, then major components of shape variation in nature may represent the consequences of developmental instability and epistasis, sources of variance which cannot be controlled via inheritance. This can be understood as the opposite of canalization – aspects of development which are less canalized than others for reasons intrinsic to the developmental system. Stabilizing selection would be less able to reduce variation in these directions than in more additively controlled/("canalized"?) ones, hence the large amount of variance explained by the former. There would be no reason to assume that they correspond to preferred directions of evolutionary divergence, especially as stabilizing selection should disfavour change in these directions. To the contrary, evolutionary divergence by selection or drift should rather occur along directions of variation which can be more easily controlled via inheritance, because they are less prone to vary through developmental instability and epistasis. This scenario represents a different twist to the idea of "lines of least resistance". It is schematically represented in Fig.25.

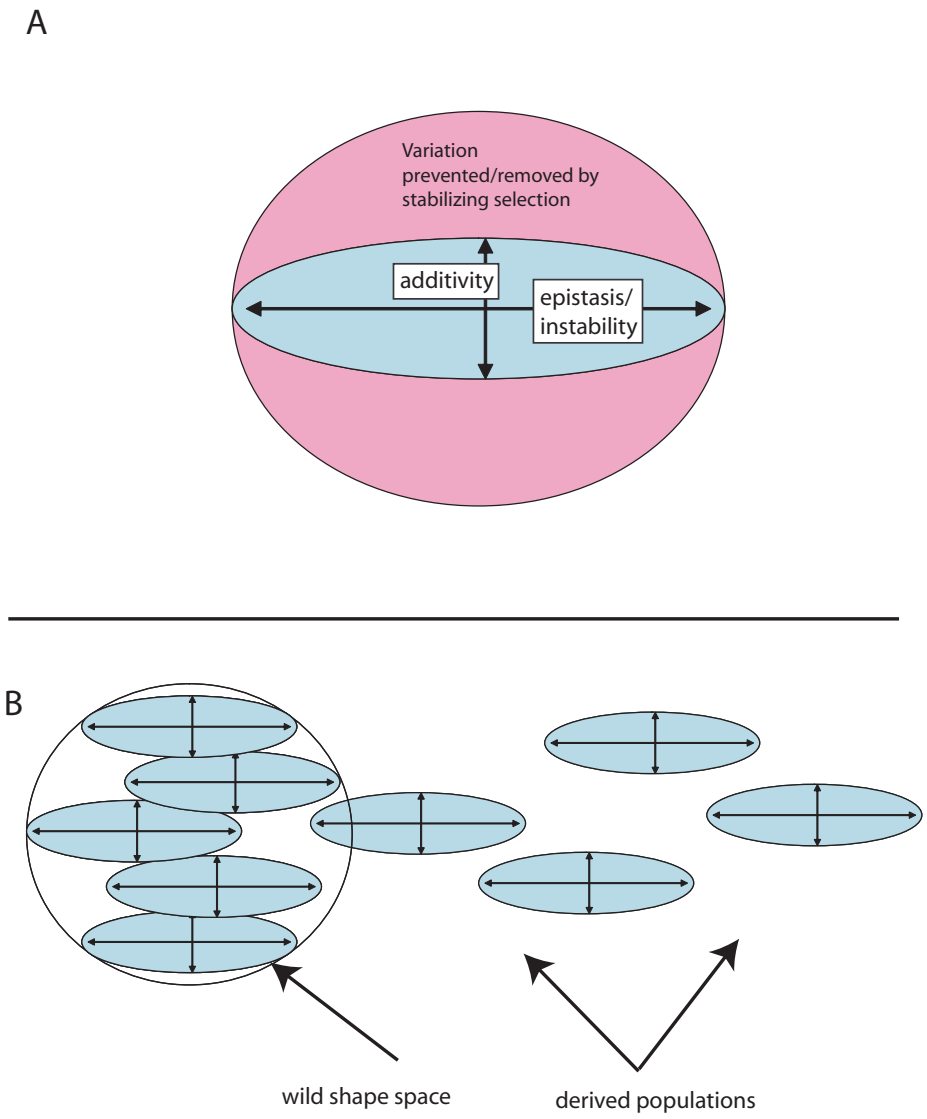


Fig.25. Diagram illustrating the concept about the relationship between additive and epistatic variance described in the text. A: within populations, the major axes of variance are dominated by epistatic genetic effects and developmental instability, which are difficult to control. Stabilizing selection can restrict the morphospace only in directions of additive genetic variance (one might consider adding associating "canalization" with additivity). B: evolutionary diversification in the wild shape space is along axes of additive genetic variation, which are minor axes of overall variation. The divergence of bottlenecked populations is along the axes of epigenetic variance, which correspond to the large axes of within-population variation. The unidirectionality of the divergence may result from developmental constraints.

This scenario, although admittedly speculative, explains all features of the observed shape space: the limited number of directions of epistatic change through inbreeding, the concordance of these directions with the major directions of variance within wild populations, and the discordance of the directions of intra- and interspecific variation with these directions.

It could be tested by comparing developmental instability and heritability for the 5 traits to unrelated directions (e.g., the directions of interpopulation or interspecific divergence). Heritability should be lower and instability should be higher in the directions of the 5 traits.

SUPPLEMENT 1

Collection numbers of museum material and museum addresses

DOM EGYPT
(ZFMK)

99367
99436
99437
99439
99443
99446
99468
99472
99477
99480
99488
99497
99501
99502
99507
99509
99514
99522
99523
99528
99532
99533
99534
99537
99543
99548
99549
99553
99558

MAC GREECE
(Senckenberg)

44931
44932
44935
44936
44937
44938
44939
44942
44943
44946
44947
44948
44951
44952
44953

DOM GER FRANKFURT
(Senckenberg)

9321
9361
9389
9390
9392
9396
9466
9527
9793
9895
9932
9948
9970
9987
10018
10020
10021
10023
10029
10263
45123
45124
45159

MAC TURKEY
(Senckenberg)

36663
36664
36665
36666
36669
36670
36671
36672
36673
36674
36675
36677
36678
36679
36681
36682
36684

DOM IRAN TEHERAN
(Senckenberg)

46375
46376
46377
46378
46379
46381
46382
46383
46385
46386
46387
46388
46389
46390
46391
46392

DOM SPAIN PUDEMONT
(Senckenberg)

31614
31616
31617
31622
31623
31624
31625
31626
31630
31632
31633
31634
31635
31636
31649
31650
31658
31659
31660
31661
31662
31664
31665
31666
31667
31749

SUPPLEMENT 1 (continued)

DOM SICILY (Senckenberg)	SPR SPAIN PUDEMONT (Senckenberg)	MUS HUNGARY (Senckenberg)
17332	31516	52150
17333	31612	52151
17335	31613	52152
17336	31615	52153
17337	31618	52154
17338	31619	52155
17339	31620	52156
17340	31621	52157
17341	31627	52158
17343	31629	52159
17344	31631	52160
17347	31637	52161
17348	31639	58228
17349	31640	58230
17350	31644	58231
17351	31645	58233
17352	31646	58234
17353	31647	58234
17354	31651	58235
17355	31652	58236
17356	31654	58238
17357	31656	58239
17358	31657	58241
17361	31744	58242
17362	31745	58243
17363	31746	
17364	31747	
37020	31748	
37022		

CAS JOHNSTON ATOLL (Smithsonian)	DOM GER MUNICH (ZSM)
360999	All specimens 1977 Kleinlangenheim,
361001	no individual numbers available.
361002	
361003	
361004	
361005	
361008	
361009	
361010	
362132	
362135	
362137	
362141	
362143	

Adresses of the museums

Senckenberg:

Senckenberganlage 5
D-60325 Frankfurt am Main
Germany

ZSM:

Zoologische Staatssammlung München
Münchhausenstrasse 21
D-81247 München
Germany

ZFMK:

Zoologisches Forschungsmuseum Alexander König
Adenauerallee 160
D53113 Bonn
Germany

Smithsonian:

Smithsonian Institution
National Museum Of Natural History
Department of Vertebrate Zoology
Division of Mammals
NHB MRC 108; 10th & Constitution Ave,N.W.
Box 37012
Washington, DC 20013-7012
USA

SUPPLEMENT 2

Anatomical definition of landmarks

LM	Anatomical definition
1	Anterior terminus of bone dorsal of the incisor
2	Minimum of depression on dorsal side of incisor ramus
3	Bone/teeth transition anterior of M1
4	Intersection of ascending ramus with tooth row
5	Tip of processus coronoideus
6	Minimum of depression posterior to processus coronoideus
7	Anterior margin of condylar articular surface
8	Posteroventral tip of condyle
9	Minimum of depression formed by condyle and processus angularis
10	Posterodorsal tip of processus angularis
11	Posteroventral tip of processus angularis
12	Minimum of depression formed by processus angularis and incisor ramus
13	Posterior margin of muscle insertion area on ventral side of incisor ramus
14	Anterior margin of muscle insertion area on ventral site of incisor ramus

SUPPLEMENT 3

R Function to average raw coordinates from 2 or 3 rounds of landmark digitization

Input (x): raw coordinate matrix. The columns contain the coordinates (x1 y1 x2 y2 etc), the rows represent the specimens, appended 2 or 3 times (specimens have to be in the same order in each block).
Output (f): a matrix of averaged raw coordinates.

2 x digitization

```
average<-function(x){
a<-data.matrix(x)
b<-a[1:(dim(a)[1]/2),]
c<-a[((dim(a)[1]/2)+1):(dim(a)[1]),]
e<-b+c
f<-e/2
write.table(f,"clipboard-256",sep="\t",col.names=NA)}
```

3 x digitization

```
average<-function(x){
a<-data.matrix(x)
b<-a[1:(dim(a)[1]/3),]
c<-a[((dim(a)[1]/3)+1):(dim(a)[1]*2/3),]
d<-a[((dim(a)[1]*2/3)+1):dim(a)[1],]
e<-b+c+d
f<-e/3
write.table(f,"clipboard-256",sep="\t",col.names=NA)}
```

SUPPLEMENT 4

R script to transform a triangular matrix into a list

Comment: I needed this script to transform the Mahalanobis distances obtained in MorphoJ into a list format making it possible to sort them manually into categories. Mahalanobis distances are exported from MorphoJ by copy and paste into a txt file. After replacing the commas from MorphoJ by dots and inserting an "x" as entry for the missing element [1,1], the file containing the triangular matrix of Mahalanobis distances is saved as "xxx.txt" The script uses a specialized function from the library "reshape" to "melt" the triangular matrix into a rectangular matrix. Unnecessary rows containing NAs resulting from the empty upper triangle are removed using "remove.NA" from the library "agce".

```
y<-read.table("xxx.txt",fill= TRUE, header = TRUE, as.is = TRUE)
row.names(y)<-y[,1]
y<-y[,-1]
z<-as.matrix(y)
library(reshape)
a<-melt(z)
library(agce)
b<-remove.NA(a)
write.table(b, "clipboard", sep = "\t")
```

R functions to bootstrap the correlations between ILMD vectors.

Comment: I wanted to know whether the frequently high correlations observed between ILMD vectors were due to a few outlier ILMDs. Therefore, I wrote this function to construct 10000 matrices containing resampled ILMD vectors (sampled from the original vectors with replacement) of the same length as the original vectors, calculate the correlation coefficients between all ILMD vectors within each of these matrices, and obtain summary statistics for the entire bootstrap.

Internal function: corvec

The internal function calculates one single bootstrap round.

Argument (x): a matrix (rows=ILMD differences, columns = taxa)
Output (c): a one column matrix containing the Pearson correlation coefficients between the columns; the row names indicate to which combination of populations each coefficient belongs.

```
corvec<-function(x) {
a<-cor(x)                                   # cor(x) produces a correlation matrix
b<-a[lower.tri(a)]                          # lower.tri selects the lower triangle
Profrom=matrix(c(rep(rownames(a),ncol(a))),nrow(a),ncol(a))
From<-c(t(Profrom))
To<-c(Profrom)
ProFromTo<-paste(From,To)
FromTo<-ProFromTo[lower.tri(a)]             #FromTo is a vector giving all pairwise
c<-matrix(b)                                #combinations of population names
rownames(c)<-FromTo
return(c)
}
```

External function: ILMDBS

The external function calculates 10000 bootstrap rounds and gives summary statistics (minimum, maximum, average and standard deviation) of the correlation coefficients for each combination of populations.

Input (x): a matrix (rows=ILMD differences, columns = taxa)
Output (p): matrix of summary statistics of the obtained correlation coefficients

```
ILMDBS<-function(x) {
mat<-c(1:length(x))
for (i in 1:10000) {
a<-sample(1:nrow(x),replace=TRUE)
BS<-x[a,]                                   #The loop is the bootstrap proper. Within each
BScor<-corvec(BS)                           # round, all columns are being sampled simul-
```

SUPPLEMENT 5 (continued)

```
mat<-cbind(mat,BScor)                    # taneously by vector a and the correlation
}                                        #coefficients are calculated and stored in mat.
as.matrix(mat)
matr<-mat[,-1]
Mean<-apply(matr,1,mean)
SD<-apply(matr,1,sd)
Min<-apply(matr,1,min)
Max<-apply(matr,1,max)
Measured<-corvec(x)
p<-cbind(Measured,Mean,SD,Min,Max)
colnames(p)<-c("Measured","BS Mean", "BS SD", "BS Min", "BS Max")
write.table(p, "clipboard", sep = "\t")
return(p)
}
```

SUPPLEMENT 6

R functions for counting the number of elements shared between all pairwise combinations of columns in a matrix.

Comment: these functions were needed because I did not want to determine manually how many ILMDs were shared among the 10 largest positive and negative ILMDs (see chapter 7) for each pair of derived populations. The internal function (count) determines for a given derived population how many ILMDs it shares with each of the other derived populations. The external function (countn) automates this for the entire set of derived populations.

Internal function: count

Argument (x):	a vector of the 10 largest positive or negative ILMDs of the derived populations.
Argument (a):	a matrix whose columns are the derived populations and whose rows are the 10 largest positive or negative ILMDs.
Output (f):	a vector of 21 places (number of derived populations) indicating how many ILMDs are shared the between derived population x and each of the derived populations (including x, the corresponding element has always a value of 10).

```
count<-function(x,a){
n<-ncol(a)
f<-0
for (i in 1:n){
inlen<-length(intersect(a[,x],a[,i]))
f<-c(f,inlen)
}
return(f)
}
```

External function: countn

Argument (a):	a matrix whose columns are the derived populations and whose rows are the 10 largest positive or negative ILMDs.
Output (names):	a square matrix of the derived populations whose entries are the number of ILMDs shared between pairs of populations.

```
countn<-function(a){
names<-c(1:ncol(a))
for (u in 1:ncol(a)){
names<-cbind(names, count(u))
}
write.table(names,"clipboard",sep = "\t")
}
```

SUPPLEMENT 7

Correlation coefficients between the ILMDd vectors of the derived populations

The lower triangle contains linear correlation coefficients, the upper triangle contains p values.

	Dom Kerg Gouillou	Dom Kerg Cochons	Cas Wd Inbred Cast/EiJ	Dom Wd Inbred Stlt	Dom Wd Inbred StrA	Dom Wd Inbred StrB	Mus Wd Inbred PWD	Dom Lab Iran Ahvaz Gen3	Dom Lab Ger Cologne Gen3	Mus Lab Khaz Gen1
Dom Kerg Gouillou	x	0.000	0.000	0.000	0.000	0.013	0.000	0.786	0.000	0.000
Dom Kerg Cochons	0.743	x	0.000	0.000	0.000	0.443	0.000	0.325	0.496	0.000
Cas Wd Inbred Cast/EiJ	0.713	0.382	x	0.000	0.000	0.000	0.000	0.133	0.000	0.000
Dom Wd Inbred Stlt	0.707	0.536	0.556	x	0.000	0.004	0.000	0.942	0.005	0.000
Dom Wd Inbred StrA	0.624	0.377	0.528	0.835	x	0.000	0.000	0.497	0.001	0.000
Dom Wd Inbred StrB	0.260	-0.081	0.483	0.302	0.453	x	0.000	0.001	0.000	0.000
Mus Wd Inbred PWD	0.604	0.416	0.698	0.411	0.553	0.381	x	0.411	0.001	0.001
Dom Lab Iran Ahvaz Gen3	0.029	0.104	0.159	0.008	0.072	0.343	0.087	x	0.000	0.000
Dom Lab Ger Cologne Gen3	0.420	0.072	0.557	0.295	0.329	0.571	0.337	0.691	x	0.000
Mus Lab Khaz Gen1	0.510	0.504	0.588	0.486	0.513	0.417	0.355	0.441	0.439	x
Spr Wd (Inbred) Spain Madrid	-0.025	0.022	0.096	-0.060	-0.196	0.120	-0.085	0.657	0.646	0.125
Cas Wd (Inbred) Taiwan	0.167	0.119	0.295	-0.085	-0.131	0.153	0.176	0.737	0.760	0.190
Dom Lab Iran Ahvaz Gen1	0.109	0.248	0.128	0.133	0.287	0.069	0.313	0.630	0.208	0.489
Dom Lab France Gen3/4	0.554	0.264	0.587	0.346	0.380	0.525	0.400	0.685	0.910	0.569
BALB/cByJ	0.436	0.598	0.169	0.407	0.327	-0.209	0.340	-0.148	-0.147	0.069
FVB/NJ	0.606	0.545	0.488	0.480	0.437	-0.103	0.536	-0.037	0.109	0.150
C57BL/10J	0.364	0.661	-0.086	0.392	0.254	-0.211	0.031	-0.068	-0.115	0.130
C57BL/6J	0.301	0.414	-0.005	0.330	0.240	-0.181	0.192	0.067	-0.028	0.051
C57BL/6J x PWD	0.272	0.331	0.007	0.296	0.265	-0.012	0.164	0.476	0.301	0.176
C57BL/6J x Cast/EiJ	-0.106	-0.059	0.086	-0.093	-0.056	0.270	-0.040	0.859	0.634	0.250
France x Iran	0.234	0.018	0.398	0.016	0.145	0.452	0.306	0.812	0.809	0.464

SUPPLEMENT 7 (continued)

	Spr Wd (Inbred) Spain Madrid	Cas Wd (Inbred) Taiwan	Dom Lab Iran Ahvaz Gen1	Dom Lab France Gen3/4	BALB/cByJ	FVB/NJ	C57BL/10J	C57BL/6J	C57BL/6J x PWD	C57BL/6J x Cast/EiJ	France x Iran
Dom Kerg Gouillou	0.813	0.114	0.305	0.000	0.000	0.000	0.000	0.004	0.009	0.320	0.026
Dom Kerg Cochons	0.839	0.260	0.018	0.011	0.000	0.000	0.000	0.000	0.001	0.580	0.864
Cas Wd Inbred Cast/EiJ	0.363	0.005	0.226	0.000	0.110	0.000	0.419	0.965	0.948	0.419	0.000
Dom Wd Inbred Stlt	0.574	0.426	0.210	0.001	0.000	0.000	0.000	0.001	0.004	0.382	0.877
Dom Wd Inbred StrA	0.063	0.215	0.006	0.000	0.002	0.000	0.015	0.022	0.011	0.599	0.170
Dom Wd Inbred StrB	0.256	0.147	0.516	0.000	0.047	0.333	0.045	0.085	0.912	0.010	0.000
Mus Wd Inbred PWD	0.423	0.095	0.002	0.000	0.001	0.000	0.774	0.069	0.120	0.709	0.003
Dom Lab Iran Ahvaz Gen3	0.000	0.000	0.000	0.000	0.163	0.727	0.522	0.525	0.000	0.000	0.000
Dom Lab Ger Cologne Gen3	0.000	0.000	0.048	0.000	0.164	0.305	0.279	0.794	0.004	0.000	0.000
Mus Lab Khaz Gen1	0.237	0.072	0.000	0.000	0.518	0.157	0.218	0.628	0.095	0.017	0.000
Spr Wd (Inbred) Spain Madrid	x	0.000	0.742	0.000	0.026	0.440	0.216	0.268	0.090	0.000	0.000
Cas Wd (Inbred) Taiwan	0.839	x	0.040	0.000	0.718	0.081	0.453	0.667	0.000	0.000	0.000
Dom Lab Iran Ahvaz Gen1	0.035	0.215	x	0.000	0.133	0.048	0.462	0.000	0.000	0.000	0.000
Dom Lab France Gen3/4	0.555	0.713	0.370	x	0.722	0.004	0.837	0.142	0.000	0.000	0.000
BALB/cByJ	-0.233	0.038	0.159	-0.038	x	0.000	0.000	0.000	0.000	0.944	0.336
FVB/NJ	-0.082	0.184	0.208	0.298	0.654	x	0.000	0.000	0.000	0.682	0.596
C57BL/10J	-0.131	-0.080	0.078	0.022	0.521	0.448	x	0.000	0.000	0.063	0.021
C57BL/6J	-0.117	0.046	0.419	0.155	0.493	0.655	0.723	x	0.000	0.789	0.678
C57BL/6J x PWD	0.179	0.385	0.581	0.389	0.446	0.495	0.509	0.818	x	0.000	0.000
C57BL/6J x Cast/EiJ	0.733	0.807	0.397	0.574	-0.007	-0.043	-0.195	-0.028	0.444	x	0.000
France x Iran	0.558	0.768	0.568	0.856	-0.102	0.056	-0.242	0.044	0.403	0.765	x

SUPPLEMENT 8

ILMDs most changed in the derived populations respective to base populations

Dom Kerg Gouillou	Dom Kerg Cochons	Cas Wd Inbred Cast/EiJ	Dom Wd Inbred Stlt	Dom Wd Inbred StrA	Dom Wd Inbred StrB	Mus Wd Inbred PWD	Dom Lab Iran Ahvaz Gen3	Dom Lab Ger Cologne Gen3	Mus Lab Khaz Gen1	Spr Wd (Inbred) Spain Madrid	Cas Wd (Inbred) Taiwan	Dom Lab Iran Ahvaz Gen1	Dom Lab France Gen3/4	BALB/cByJ	FVB/NJ	C57BL/10J	C57BL/6J	C57BL/6J x PWD	C57BL/6J x Cast/EiJ	France x Iran	
5-6	5-6	5-6	10-11	10-11	6-7	5-6	3-4	8-9	8-9	8-9	8-9	3-4	8-9	1-14	5-6	10-11	12-13	3-4	8-10	3-4	10
11-12	7-9	8-9	5-6	5-6	8-10	3-4	8-10	6-7	7-9	8-10	8-10	12-13	8-10	2-13	4-5	1-14	4-6	2-4	8-9	8-9	ILMD
8-10	3-4	11-12	8-11	10-12	11-12	6-7	2-4	8-10	13-14	7-9	7-9	2-4	11-12	5-6	4-6	1-3	4-5	1-14	3-4	6-7	most
8-9	8-10	6-7	9-11	3-4	8-9	3-5	7-9	7-12	1-3	7-10	3-4	2-12	3-4	1-13	3-5	7-8	9-11	2-12	13-14	11-12	increased
3-4	14	8-10	3-4	11-12	10-12	2-4	1-4	3-4	11-14	8-11	7-10	12-14	11-13	2-14	2-5	12-13	3-5	2-6	2-13	8-10	in
9-11	2-4	3-4	8-10	8-9	8-11	11-12	8-9	8-12	5-6	9-10	9-10	2-13	7-12	3-5	3-6	1-2	1-14	3-6	2-4	11-13	derived
8-11	7-10	8-12	8-9	9-12	9-10	2-5	3-7	11-12	2-3	3-4	7-8	13-14	8-12	3-13	1-5	2-3	3-6	2-13	6-7	11-14	populations
7-11	6-9	4-5	9-10	8-12	11-13	6-8	2-7	9-10	12-14	7-8	8-12	1-4	9-10	1-5	2-6	4-6	2-5	2-3	7-9	3-7	
7-9	1-2	3-5	7-11	8-11	5-7	2-13	3-6	7-13	11-13	7-11	2-4	1-12	6-7	4-5	5-12	1-4	2-12	4-6	8-11	13-14	
8-12	1-3	7-12	8-12	2-4	4-7	4-5	7-12	3-7	1-2	5-10	7-12	1-13	9-11	10-11	6-12	4-5	2-6	2-5	1-13	2-4	
10-11	4-12	3-12	4-9	5-10	4-14	12-14	5-12	2-13	1-14	6-7	10-14	10-12	6-9	9-14	4-13	7-14	3-14	5-11	5-9	2-3	10
2-12	6-8	1-14	4-12	2-14	6-14	13-14	4-12	5-7	3-14	4-5	3-14	4-10	2-13	11-14	6-8	4-14	10-12	1-2	5-13	4-10	ILMD
6-9	12-14	4-9	3-13	12-13	4-12	4-6	4-10	4-12	5-9	1-3	5-7	5-11	3-13	10-13	12-14	4-7	8-9	11-12	10-12	4-9	most
2-13	4-8	12-13	12-13	2-3	5-9	2-14	7-8	6-9	7-8	10-12	10-12	5-9	5-8	11-13	1-2	4-12	11-12	3-14	3-14	4-5	decreased
12-14	4-6	7-8	6-9	3-13	4-6	2-3	3-13	4-9	5-10	1-14	12-13	5-10	3-14	9-13	3-14	12-14	5-9	13-14	10-13	3-14	in
4-6	13-14	5-7	2-14	1-2	2-14	1-2	4-5	3-14	3-13	2-14	10-13	5-8	2-3	5-8	10-11	5-8	6-7	10-11	2-14	6-9	derived
13-14	4-7	5-8	5-9	5-8	3-13	3-13	3-14	5-6	4-6	4-9	2-14	9-10	5-9	6-7	12-13	3-13	7-9	5-8	4-5	2-14	populations
5-9	5-8	4-12	3-14	5-9	5-8	5-7	2-14	5-9	5-8	9-12	4-9	10-11	2-14	12-14	6-7	6-7	5-8	5-6	12-13	5-6	
5-8	6-7	10-11	5-8	3-14	3-14	10-11	10-11	2-14	2-14	5-6	5-6	3-14	5-7	12-13	5-8	13-14	13-14	5-9	10-11	5-9	
5-7	5-7	2-14	5-7	5-7	7-8	3-14	5-6	10-11	5-7	10-11	10-11	5-7	10-11	5-7	5-7	5-7	5-7	5-7	5-6	10-11	

Correlation coefficients between preliminary traits.

The lower triangle contains linear correlation coefficients, the upper triangle contains p values.

	5-6-plus	11-12-plus	8-10-plus	8-9-plus	3-4-plus	8-12-plus	2-4-plus	4-5-plus	9-10-plus	1-14-plus	4-6-plus	10-11-min	6-9-min	12-14-min	5-9-min	5-8-min	5-7-min	6-7-min	2-14-min	3-14-min
5-6-plus	x	0.66	0.72	0.62	0.27	0.61	0.00	0.00	0.88	0.97	0.65	0.72	0.37	0.00	0.30	0.00	0.00	0.06	0.46	0.53
11-12-plus	0.09	x	0.00	0.00	0.00	0.00	0.15	0.12	0.00	0.00	0.00	0.11	0.00	0.24	0.00	0.14	0.06	0.01	0.00	0.00
8-10-plus	-0.08	0.86	x	0.00	0.00	0.00	0.00	0.22	0.00	0.01	0.02	0.00	0.00	0.17	0.00	0.18	0.07	0.03	0.00	0.00
8-9-plus	0.11	0.90	0.85	x	0.00	0.00	0.14	0.70	0.00	0.03	0.06	0.00	0.00	0.72	0.00	0.04	0.02	0.25	0.00	0.00
3-4-plus	0.23	0.71	0.87	0.83	x	0.00	0.12	0.47	0.00	0.31	0.60	0.00	0.00	0.58	0.00	0.00	0.00	0.80	0.00	0.00
8-12-plus	0.11	0.88	0.83	0.95	0.77	x	0.25	0.59	0.00	0.02	0.05	0.05	0.00	0.66	0.00	0.09	0.04	0.18	0.00	0.00
2-4-plus	-0.61	0.30	0.58	0.31	0.32	0.24	x	0.06	0.16	0.27	0.26	0.00	0.70	0.01	0.06	0.27	0.67	0.09	0.05	0.02
4-5-plus	0.63	-0.32	-0.26	-0.08	0.15	-0.11	-0.38	x	0.26	0.00	0.00	0.20	0.83	0.00	0.98	0.00	0.01	0.00	0.37	0.41
9-10-plus	0.03	0.83	0.85	0.77	0.74	0.85	0.30	-0.24	x	0.01	0.02	0.16	0.00	0.33	0.00	0.36	0.17	0.05	0.00	0.00
1-14-plus	0.01	-0.75	-0.54	-0.45	-0.22	-0.47	-0.24	0.68	-0.55	x	0.00	0.74	0.16	0.01	0.02	0.89	0.62	0.00	0.00	0.59
4-6-plus	0.10	-0.69	-0.46	-0.39	-0.11	-0.41	-0.24	0.73	-0.49	0.99	x	0.49	0.27	0.00	0.08	0.66	0.93	0.00	0.00	0.92
10-11-min	0.08	0.34	0.58	0.56	0.70	0.41	0.62	0.27	0.29	0.07	0.15	x	0.04	0.50	0.00	0.01	0.00	0.48	0.01	0.00
6-9-min	0.19	0.76	0.67	0.93	0.70	0.94	0.08	0.05	0.66	-0.29	-0.23	0.41	x	0.78	0.00	0.03	0.02	0.66	0.00	0.00
12-14-min	0.80	-0.25	-0.29	-0.08	0.12	-0.09	-0.54	0.95	-0.21	0.51	0.56	0.14	0.06	x	0.96	0.00	0.00	0.00	0.52	0.64
5-9-min	0.22	0.85	0.90	0.85	0.91	0.81	0.39	0.01	0.83	-0.46	-0.37	0.58	0.67	0.01	x	0.02	0.00	0.28	0.00	0.00
5-8-min	0.86	0.31	0.28	0.42	0.57	0.35	-0.23	0.57	0.20	-0.03	0.09	0.52	0.44	0.65	0.48	x	0.00	0.19	0.03	0.01
5-7-min	0.81	0.39	0.38	0.49	0.63	0.42	-0.09	0.52	0.29	-0.11	0.02	0.59	0.47	0.60	0.59	0.98	x	0.30	0.01	0.00
6-7-min	0.39	-0.53	-0.44	-0.24	-0.05	-0.28	-0.35	0.92	-0.40	0.82	0.81	0.15	-0.10	0.85	-0.23	0.28	0.22	x	0.04	0.98
2-14-min	0.16	0.92	0.92	0.85	0.82	0.80	0.40	-0.19	0.80	-0.65	-0.57	0.50	0.67	-0.14	0.92	0.44	0.52	-0.42	x	0.00
3-14-min	0.14	0.60	0.80	0.78	0.93	0.69	0.48	0.18	0.63	-0.12	-0.02	0.83	0.64	0.10	0.86	0.52	0.61	0.01	0.72	x

SUPPLEMENT 10

Correlation coefficients between the trait score vectors of the derived populations

The lower triangle contains linear correlation coefficients, the upper triangle contains p values.

	Dom Kerg Gouillou	Dom Kerg Cochons	Cas Wd Inbred Cast/EiJ	Dom Wd Inbred Stlt	Dom Wd Inbred StrA	Dom Wd Inbred StrB	Mus Wd Inbred PWD	Dom Lab Iran Ahvaz Gen3	Dom Lab Ger Cologne Gen3	Mus Lab Khaz Gen1
Dom Kerg Gouillou	x	0.101	0.034	0.047	0.082	0.482	0.178	0.007	0.157	0.219
Dom Kerg Cochons	0.804	x	0.200	0.357	0.467	0.057	0.071	0.044	0.001	0.657
Cas Wd Inbred Cast/EiJ	0.907	0.687	x	0.029	0.016	0.671	0.102	0.107	0.284	0.046
Dom Wd Inbred Stlt	0.883	0.531	0.916	x	0.016	0.949	0.372	0.149	0.490	0.068
Dom Wd Inbred StrA	0.829	0.433	0.944	0.943	x	0.927	0.299	0.222	0.592	0.008
Dom Wd Inbred StrB	-0.420	-0.867	-0.261	-0.040	0.058	x	0.208	0.284	0.026	0.756
Mus Wd Inbred PWD	0.711	0.846	0.803	0.518	0.586	-0.678	x	0.175	0.078	0.337
Dom Lab Iran Ahvaz Gen3	-0.967	-0.889	-0.797	-0.744	-0.664	0.600	-0.714	x	0.070	0.428
Dom Lab Ger Cologne Gen3	-0.735	-0.990	-0.601	-0.412	-0.327	0.922	-0.836	0.848	x	0.786
Mus Lab Khaz Gen1	0.667	0.273	0.885	0.851	0.965	0.193	0.550	-0.467	-0.169	x
Spr Wd (Inbred) Spain Madrid	-0.955	-0.912	-0.867	-0.733	-0.712	0.630	-0.855	0.971	0.877	-0.560
Cas Wd (Inbred) Taiwan	-0.956	-0.929	-0.894	-0.797	-0.731	0.623	-0.846	0.955	0.876	-0.583
Dom Lab Iran Ahvaz Gen1	-0.605	-0.362	-0.267	-0.514	-0.305	0.133	0.069	0.647	0.310	-0.062
Dom Lab France Gen3/4	-0.528	-0.895	-0.327	-0.119	-0.032	0.981	-0.689	0.707	0.947	0.132
BALB/cByJ	0.638	0.844	0.339	0.224	0.117	-0.854	0.548	-0.813	-0.882	-0.105
FVB/NJ	0.759	0.965	0.552	0.404	0.303	-0.895	0.744	-0.890	-0.979	0.109
C57BL/10J	0.189	0.466	-0.210	-0.184	-0.379	-0.663	0.003	-0.417	-0.530	-0.603
C57BL/6J	0.236	0.639	-0.089	-0.124	-0.339	-0.841	0.213	-0.461	-0.697	-0.539
C57BL/6J x PWD	-0.914	-0.665	-0.991	-0.892	-0.950	0.242	-0.796	0.809	0.586	-0.886
C57BL/6J x Cast/EiJ	-0.964	-0.924	-0.833	-0.741	-0.678	0.643	-0.795	0.992	0.885	-0.500
France x Iran	-0.767	-0.953	-0.530	-0.445	-0.294	0.864	-0.662	0.895	0.954	-0.080

SUPPLEMENT 10 (continued)

	Spr Wd (Inbred) Spain Madrid	Cas Wd (Inbred) Taiwan	Dom Lab Iran Ahvaz Gen1	Dom Lab France Gen3/4	BALB/cByJ	FVB/NJ	C57BL/10J	C57BL/6J	C57BL/6J x PWD	C57BL/6J x Cast/EiJ	France x Iran
Dom Kerg Gouillou	0.012	0.011	0.280	0.360	0.246	0.136	0.761	0.702	0.030	0.008	0.130
Dom Kerg Cochons	0.031	0.022	0.549	0.040	0.072	0.008	0.429	0.245	0.221	0.025	0.012
Cas Wd Inbred Cast/EiJ	0.057	0.041	0.665	0.591	0.577	0.335	0.735	0.887	0.001	0.080	0.358
Dom Wd Inbred Stlt	0.159	0.106	0.376	0.849	0.717	0.500	0.767	0.843	0.042	0.152	0.452
Dom Wd Inbred StrA	0.177	0.161	0.618	0.959	0.852	0.620	0.529	0.577	0.013	0.209	0.631
Dom Wd Inbred StrB	0.255	0.262	0.832	0.003	0.066	0.040	0.222	0.074	0.695	0.242	0.059
Mus Wd Inbred PWD	0.065	0.071	0.912	0.198	0.339	0.150	0.996	0.731	0.107	0.108	0.223
Dom Lab Iran Ahvaz Gen3	0.006	0.011	0.238	0.181	0.095	0.043	0.485	0.435	0.098	0.001	0.040
Dom Lab Ger Cologne Gen3	0.051	0.052	0.612	0.015	0.048	0.004	0.358	0.190	0.299	0.046	0.012
Mus Lab Khaz Gen1	0.327	0.302	0.921	0.832	0.867	0.862	0.282	0.348	0.045	0.391	0.899
Spr Wd (Inbred) Spain Madrid	x	0.004	0.448	0.170	0.136	0.049	0.655	0.546	0.049	0.001	0.067
Cas Wd (Inbred) Taiwan	0.977	x	0.445	0.201	0.189	0.063	0.709	0.543	0.050	0.003	0.066
Dom Lab Iran Ahvaz Gen1	0.449	0.452	x	0.670	0.333	0.432	0.294	0.453	0.648	0.337	0.318
Dom Lab France Gen3/4	0.720	0.686	0.262	x	0.020	0.013	0.177	0.075	0.591	0.157	0.028
BALB/cByJ	-0.760	-0.699	-0.554	-0.934	x	0.012	0.091	0.077	0.546	0.105	0.015
FVB/NJ	-0.880	-0.857	-0.463	-0.951	0.955	x	0.236	0.141	0.333	0.036	0.002
C57BL/10J	-0.275	-0.231	-0.591	-0.712	0.818	0.649	x	0.018	0.754	0.549	0.195
C57BL/6J	-0.365	-0.367	-0.445	-0.840	0.838	0.754	0.939	x	0.862	0.455	0.110
C57BL/6J x PWD	0.880	0.879	0.280	0.327	-0.365	-0.554	0.194	0.109	x	0.076	0.370
C57BL/6J x Cast/EiJ	0.991	0.980	0.550	0.735	-0.798	-0.903	-0.362	-0.443	0.839	x	0.040
France x Iran	0.852	0.853	0.568	0.918	-0.945	-0.986	-0.693	-0.793	0.519	0.896	x

SUPPLEMENT 11

Correlation coefficients between preliminary traits in wild *Mus musculus (without the Kerguelen populations)*

The lower triangle contains linear correlation coefficients, the upper triangle contains p values.

	5-6-plus	11-12-plus	8-10-plus	8-9-plus	3-4-plus	8-12-plus	2-4-plus	4-5-plus	9-10-plus	1-14-plus	4-6-plus	10-11-min	6-9-min	12-14-min	5-9-min	5-8-min	5-7-min	6-7-min	2-14-min	3-14-min
5-6-plus	x	0.39	0.53	0.00	0.00	0.00	0.00	0.00	0.04	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.10	0.00
11-12-plus	0.06	x	0.00	0.00	0.00	0.00	0.74	0.23	0.00	0.00	0.00	0.00	0.00	0.83	0.00	0.15	0.12	0.02	0.00	0.01
8-10-plus	0.05	0.70	x	0.00	0.00	0.00	0.00	0.78	0.00	0.00	0.00	0.00	0.00	0.85	0.00	0.00	0.00	0.12	0.00	0.00
8-9-plus	0.22	0.85	0.74	x	0.00	0.00	0.69	0.18	0.00	0.00	0.03	0.00	0.00	0.08	0.00	0.00	0.00	0.91	0.00	0.00
3-4-plus	0.54	0.37	0.71	0.62	x	0.00	0.45	0.00	0.00	0.00	0.00	0.62	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
8-12-plus	0.22	0.76	0.66	0.84	0.50	x	0.03	0.07	0.00	0.01	0.03	0.00	0.00	0.01	0.00	0.00	0.00	0.28	0.00	0.00
2-4-plus	-0.64	0.02	0.31	0.03	-0.05	-0.16	x	0.00	0.92	0.01	0.00	0.01	0.01	0.00	0.40	0.00	0.00	0.00	0.00	0.08
4-5-plus	0.72	-0.09	-0.02	0.10	0.47	0.13	-0.50	x	0.11	0.00	0.00	0.00	0.13	0.00	0.00	0.00	0.00	0.00	0.02	0.00
9-10-plus	0.15	0.61	0.66	0.66	0.50	0.82	-0.01	0.11	x	0.02	0.05	0.00	0.00	0.05	0.00	0.05	0.02	0.24	0.00	0.00
1-14-plus	0.28	-0.48	-0.27	-0.21	0.27	-0.20	-0.19	0.71	-0.17	x	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.25	0.00
4-6-plus	0.35	-0.43	-0.21	-0.15	0.35	-0.16	-0.20	0.73	-0.14	0.99	x	0.00	0.02	0.00	0.00	0.00	0.00	0.00	0.67	0.00
10-11-min	-0.18	0.71	0.48	0.53	-0.04	0.46	0.18	-0.60	0.37	-0.92	-0.89	0.00	0.00	0.00	0.58	0.08	0.13	0.00	0.00	0.11
6-9-min	0.23	0.72	0.55	0.85	0.41	0.92	-0.18	0.11	0.66	-0.20	-0.16	0.47	x	0.02	0.00	0.00	0.00	0.59	0.00	0.00
12-14-min	0.86	-0.02	-0.01	0.13	0.46	0.19	-0.64	0.94	0.14	0.51	0.55	-0.42	0.17	x	0.00	0.00	0.00	0.00	0.06	0.00
5-9-min	0.50	0.46	0.62	0.63	0.83	0.46	0.06	0.52	0.52	0.28	0.34	-0.04	0.36	0.48	x	0.00	0.00	0.00	0.00	0.00
5-8-min	0.86	0.10	0.28	0.39	0.73	0.23	-0.28	0.58	0.14	0.30	0.39	-0.13	0.28	0.66	0.60	x	0.00	0.00	0.00	0.00
5-7-min	0.84	0.11	0.30	0.39	0.73	0.23	-0.20	0.57	0.16	0.28	0.37	-0.11	0.26	0.64	0.66	0.99	x	0.00	0.00	0.00
6-7-min	0.52	-0.17	-0.11	-0.01	0.29	0.08	-0.43	0.92	0.08	0.67	0.65	-0.61	0.04	0.86	0.36	0.34	0.33	x	0.31	0.00
2-14-min	0.12	0.66	0.73	0.66	0.57	0.47	0.33	0.17	0.56	-0.08	-0.03	0.29	0.41	0.13	0.81	0.27	0.33	0.07	x	0.00
3-14-min	0.50	0.19	0.53	0.54	0.88	0.33	0.13	0.40	0.34	0.34	0.42	-0.11	0.30	0.38	0.82	0.76	0.80	0.22	0.53	x

SUPPLEMENT 12

Percent of variance explained by the 5 traits within populations

n.s. means non significant.

Population		T812	T810	T1011	T56	T67	Joint
Wild mice							
Dom Egypt	total	9.8	7.6	n.s.	26	22.9	54.2
	PC1	14.8	n.s.	n.s.	82.4	69.5	91.4
Dom Ger Frankfurt	total	9.9	8.5	11.1	13.4	13.9	53.9
	PC1	n.s.	n.s.	32.6	n.s.	38.5	72
Dom Iran Teheran	total	n.s.	12.8	n.s.	26	23.1	66.6
	PC1	n.s.	n.s.	n.s.	72	38.3	88.7
Dom Spain PudeMont	total	9.3	8.2	10.9	26.1	20	51.7
	PC1	n.s.	n.s.	16.5	88.4	52.3	88.7
Dom Iran Ahvaz	total	n.s.	11.7	10.8	18.4	23.5	50.2
	PC1	n.s.	22.9	n.s.	44	74.2	87.7
Dom Ger Munich	total	n.s.	n.s.	n.s.	15.1	14.3	61.8
	PC1	n.s.	n.s.	n.s.	n.s.	n.s.	72.9
Dom Sicily	total	10.7	7.5	8.5	21.6	17.7	48.5
	PC1	n.s.	n.s.	n.s.	68.6	50.7	76.9
Cas Johnston Atoll	total	14.9	n.s.	n.s.	25.1	25.4	72.2
	PC1	n.s.	n.s.	n.s.	66.3	64.8	92
Mus Hungary	total	10	8.4	n.s.	21.8	28.7	59
	PC1	n.s.	n.s.	n.s.	49.7	76.9	91.1
Mac Greece	total	n.s.	15.9	n.s.	14.7	n.s.	56.9
	PC1	n.s.	47.9	n.s.	n.s.	n.s.	79.6
Mac Turkey	total	14.1	19.3	n.s.	22.8	16.6	57.7
	PC1	n.s.	51.4	n.s.	63.3	36.6	88.1
Spr Spain PudeMont	total	8	10.4	9.2	16.6	12	49.4
	PC1	n.s.	n.s.	n.s.	30.1	n.s.	61.1
Spr Spain Madrid	total	6.9	9.5	6.7	14	11.8	42.2
	PC1	n.s.	n.s.	n.s.	31.9	27.7	44.6
Outbred captive populations							
Dom Lab France Gen3/4	total	n.s.	n.s.	26.9	26.7	36.1	75.1
	PC1	n.s.	n.s.	56.8	40.3	79.1	95.4
Dom Lab Iran Ahvaz Gen1	total	11.6	9.3	n.s.	31.9	29.4	61.1
	PC1	n.s.	n.s.	n.s.	82	70.4	92
France x Iran	total	8.8	5.9	6.8	8.5	9.4	38.8
	PC1	13.5	9	8.8	n.s.	n.s.	47.1
Dom Lab Iran Ahvaz Gen3/4	total	19.1	n.s.	n.s.	18.4	n.s.	57.5
	PC1	45.1	n.s.	n.s.	38.4	n.s.	66.5
Dom Lab Ger Cologne	total	n.s.	n.s.	n.s.	17.4	20.9	50.6
	PC1	n.s.	n.s.	n.s.	n.s.	n.s.	51.7
Mus Lab Khaz	total	n.s.	14.3	n.s.	12.7	24.6	61.9
	PC1	n.s.	n.s.	n.s.	n.s.	64.8	77.3
Spre Wd (Inbred) Spain Madrid	total	n.s.	n.s.	n.s.	25.4	16.9	55.2
	PC1	n.s.	n.s.	n.s.	52	n.s.	70.9
Cas Wd (Inbred) Taiwan	total	18.3	n.s.	20.8	27.1	21.9	70.5
	PC1	34	n.s.	45	70.4	49.9	92.5

SUPPLEMENT 12 (continued)

Population		T812	T810	T1011	T56	T67	Joint
Inbred strains and Kerguelen							
BALB/cByJ	total	n.s.	9.6	6.2	8.9	13.2	41.3
	PC1	n.s.	17.3	n.s.	n.s.	n.s.	25.6
Dom Kerg Guillou	total	22.2	10.1	n.s.	26.1	30.3	64.4
	PC1	55.8	n.s.	n.s.	61.9	80.7	93
Dom Kerg Cochons	total	9.5	7.5	7.8	8	14.6	42.8
	PC1	27.9	n.s.	23.1	n.s.	30.5	64
Cas Wd Inbred Cast/EiJ	total	n.s.	n.s.	n.s.	16.5	23.8	53.4
	PC1	n.s.	n.s.	n.s.	28	47	62.3
Mus Wd Inbred PWd	total	n.s.	11.7	n.s.	19.6	11.9	52.2
	PC1	n.s.	22.5	n.s.	58.6	n.s.	82.3
Dom Wd Inbred Stlt	total	17.5	12.5	n.s.	24.7	n.s.	58.6
	PC1	28.7	n.s.	n.s.	61.2	n.s.	90.9
Dom Wd Inbred StrA	total	n.s.	n.s.	n.s.	25.6	22.1	54.5
	PC1	n.s.	n.s.	n.s.	82.8	71.5	94.2
Dom Wd Inbred StrB	total	n.s.	11.3	10.3	12.7	13.6	42.1
	PC1	n.s.	n.s.	n.s.	n.s.	n.s.	7.9
FVB/NJ	total	9.1	7.2	9.7	8.9	15.4	45.9
	PC1	n.s.	n.s.	n.s.	n.s.	30.2	51.2
C57BL/10J	total	8.1	8.9	14.1	12.7	11.2	49
	PC1	n.s.	n.s.	31	n.s.	n.s.	68.5
C57BL/6J	total	n.s.	15.9	n.s.	n.s.	28.1	68.7
	PC1	n.s.	n.s.	n.s.	n.s.	59	75

SUPPLEMENT 13

Percent of variance explained by the 5 traits between populations and species

n.s. means non significant.

		T812	T810	T1011	T56	T67	Joint
Intraspecific							
Dom Sicily	Dom Egypt	n.s.	n.s.	n.s.	n.s.	7.6	25.3
Dom Iran Ahvaz	Dom Egypt	10.7	9.7	16.1	n.s.	11.7	48.7
Dom Sicily	Dom Iran Teheran	n.s.	n.s.	n.s.	n.s.	n.s.	12.3
Dom Iran Teheran	Dom Egypt	n.s.	n.s.	n.s.	n.s.	n.s.	6.9
Dom Ger Munich	Dom Ger Frankfurt	n.s.	n.s.	n.s.	n.s.	19.3	37.5
Dom Iran Ahvaz	Dom Iran Teheran	n.s.	14.1	21.8	n.s.	8.6	48.3
Dom Sicily	Dom Iran Ahvaz	13.1	15.4	18.2	n.s.	n.s.	47.4
Mus Hungary	Dom Iran Teheran	n.s.	n.s.	n.s.	n.s.	n.s.	7.9
Dom Ger Munich	Dom Iran Teheran	13.8	n.s.	n.s.	n.s.	n.s.	26.2
Mus Hungary	Dom Ger Frankfurt	8.1	16.0	n.s.	n.s.	n.s.	17.6
Cas Johnston Atoll	Dom Sicily	n.s.	n.s.	n.s.	n.s.	46.8	53.1
Mus Hungary	Dom Iran Ahvaz	n.s.	17.5	40.6	n.s.	10.5	55.3
Mus Hungary	Cas Johnston Atoll	n.s.	n.s.	n.s.	n.s.	12.1	23.7
Mus Hungary	Dom Ger Munich	12.1	n.s.	n.s.	n.s.	n.s.	24.2
Dom Sicily	Dom Ger Munich	n.s.	n.s.	n.s.	n.s.	n.s.	8.1
Mus Hungary	Dom Egypt	n.s.	n.s.	9.2	n.s.	n.s.	10.0
Dom Ger Munich	Dom Iran Ahvaz	35.3	n.s.	16.9	n.s.	n.s.	59.4
Dom Iran Ahvaz	Dom Spain PudeMont	25.2	n.s.	23.7	14.6	n.s.	50.3
Dom Ger Munich	Dom Spain PudeMont	n.s.	n.s.	n.s.	16.7	n.s.	40.1
Mus Hungary	Dom Sicily	n.s.	n.s.	n.s.	n.s.	n.s.	12.7
Cas Johnston Atoll	Dom Iran Teheran	n.s.	n.s.	n.s.	n.s.	15.1	18.0
Dom Ger Munich	Dom Egypt	12.8	n.s.	n.s.	n.s.	10.2	31.1
Dom Spain PudeMont	Dom Egypt	n.s.	15.6	n.s.	12.2	n.s.	44.4
Cas Johnston Atoll	Dom Ger Frankfurt	n.s.	12.9	n.s.	n.s.	21.1	36.2
Cas Johnston Atoll	Dom Egypt	n.s.	n.s.	n.s.	n.s.	26.5	31.8
Dom Iran Teheran	Dom Ger Frankfurt	n.s.	13.6	n.s.	n.s.	n.s.	19.1
Dom Ger Frankfurt	Dom Egypt	n.s.	8.3	n.s.	n.s.	n.s.	14.5
Cas Johnston Atoll	Dom Iran Ahvaz	n.s.	12.8	20.9	n.s.	40.0	62.2
Dom Spain PudeMont	Dom Ger Frankfurt	n.s.	n.s.	n.s.	11.3	n.s.	27.8
Dom Iran Ahvaz	Dom Ger Frankfurt	24.5	n.s.	20.4	n.s.	15.0	39.3
Dom Sicily	Dom Ger Frankfurt	n.s.	13.5	n.s.	n.s.	9.8	32.8
Dom Sicily	Dom Spain PudeMont	n.s.	23.1	n.s.	13.9	n.s.	43.4
Mus Hungary	Dom Spain PudeMont	7.6	23.5	n.s.	12.9	n.s.	41.0
Dom Spain PudeMont	Dom Iran Teheran	n.s.	20.8	n.s.	n.s.	n.s.	35.8
Cas Johnston Atoll	Dom Ger Munich	n.s.	n.s.	n.s.	n.s.	42.4	47.8
Cas Johnston Atoll	Dom Spain PudeMont	n.s.	18.1	n.s.	n.s.	29.9	55.5
Interspecific							
Mus musculus	*Mus macedonicus*	2.7	n.s.	15.6	n.s.	11.8	25.7
Mus musculus	*Mus spretus*	15.3	n.s.	5.4	n.s.	7.5	24.8
Mus macedonicus	*Mus spretus*	8.5	15.8	15.1	n.s.	4.9	28.9

Literature

Alvarez-Castro, J.M., Kopp, M., Hermisson, J. (2009). “Effects of epistasis and the evolution of genetic architecture: exact results for a 2-locus model.” Theor Pop Biol **75**: 109-122.

Atchley, W., Cowley, E., Eisen, E.J., Prasetyo, H., Hakins-Brown, D. (1990). "Correlated response in the developmental choreographies of the mouse mandible to selection for body composition." Evolution **44**: 669-688.

Atchley, W. (1991). "A model for development and evolution of complex morphological structures." Biol Rev **66**: 101-157.

Auffray J.-C., Alibert, P., Latieule, C. (1996). "Relative warps analysis of skull shape across the hybrid zone in the house mouse (Mus musculus) in Denmark." J Zool Lond **240**: 441-455.

Badyaev, A.V. (2005). “Stress-induced variation in evolution: from behavioural plasticity to genetic assimilation.” Proc. R. Soc. B **272**: 877-886

Barton, N.H., Keightley, P.D. (2002). “Understanding quantitative genetic variation.” Nat Rev Gen **3**, 11-21.

Berry, R.J., Jakobson, M.E. (1971). "Life and death in an island population of the house mouse." Exp Geront **6**: 187-197.

Bomford, M. (1987). "Food and reproduction of wild house mice I. Diet and breeding seasons in various habitats on irrigated cereal farms in New South Wales." Aust Wildl Res **14**: 183-196.

Burgio, G., Baylac, M., Heyer, E., Montagutelli, X. (2009). “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 **63**: 2668-2686.

Bryant, E.H., Meffert, L.M. (1990). “Multivariate phenotypic differentiation among bottleneck lines of the housefly.” Evolution **44**: 660-668.

Cheverud, J.M. (1988). “A comparison of genetic and phenotypic correlations.” Evolution **42**: 958-968.

Cheverud, J.M., Ehrich, T.H., Vaughn, T.T., Koreishi, S.F., Linsey, R.B., Pletscher, L.S. (2004). "Pleiotropic effects on mandibular morphology II: differential epistasis and genetic variation on morphological integration." J Exp Zool (Mol Dev Evol) **302B**: 424-435.

Cheverud, J.M., Vaughn, T.T., Pletscher, L.S., King-Ellison, K., Bailiff, J., Adams, E., Erickson, C., Bonislawski, A. (1999). "Epistasis and the evolution of additive genetic variance in populations that pass through a bottleneck.” Evolution **53**: 1009-1018.

Cheverud, J.M., Routman, E.J. (1996). Epistasis as a source of increased additive genetic variance at population botttlenecks”. Evolution **50**: 1042-1051.

Cheverud, J.M., Routman, E.J., Irschick, D.J. (1997). "Pleiotropic effects of individual gene loci on mandibular morphology." Evolution **51**(6): 2006-2016.

Corti, M., Rohlf, F.J. (2001). "Chromosomal speciation and phenotypic evolution in the house mouse." Biol J Linn Soc **73**: 99-112.

Cucchi, T., Vigne, J.-D., Auffray, J.-C. (2005). „First occurence of the house mouse (*Mus musculus domesticus* Schwarz & Schwarz 1943) in the Western Mediterranean: a zooarcheological revision of subfossil occurrences. Biol J Linn Soc **84**: 429-445.

Davis, S.J. (1983). "Morphometric variation of populations of House mice Mus domesticus in Britain and Faroe." J Zool Lond **199**: 521-534.

Ehrich, T.H., Vaughn, T.T., Koreishi, S.F., Linsey, R.B., Pletscher, L.S., and Cheverud , J.M. (2003). "Pleiotropic effects on mandibular morphology I. Developmental morphological integration and differential dominance." J Exp Zool (Mol Dev Evol) **296B**: 58-79.

Frazer, K.A., Eskin, E., Kang, H. M., Bogue, M.A., Hinds, D.A., Beilharz, E.J., Gupta, R.V., Montgomery, J., Morenzoni, M.M., Nilsen, G.B., Pethiyagoda, C.L., Stuve, L.L., Johnson, F.M., Daly, M.J., Wade. C.M., Cox, D.R. (2007). "A sequence-based variation map of 8.27 million SNPs in inbred mouse strains." Nature **448**: 1050-1053.

Gerasimov, S., Nikolov, H., Mihailova, V., Auffray, J.-C., Bonhomme, F. (1990). "Morphometric stepwise discriminant analysis of the five genetically determined European taxa of the genus Mus." Biol J Linn Soc **41**: 47-64.

Hammer, Ø., Harper, D.A.T., Ryan, P.D. (2001). „PAST: Palaeontological Statistics Software package for education and data analysis.“ Palaeontologia Electronica **4**: 9pp.

Hansen, T.F., Alvarez-Castro, J.M., Carter, A.J. R., Hermisson, J., Wagner, G.P. (2006). “Evolution of genetic architecture under stabilizing selection.” Evolution **60**: 1523-1536.

Hermisson, J., Hansen, T.F., Wagner, G.P. (2003). “Epistasis in polygenic traits and the evolution of genetic architecture under stabilizing selection.” Am Nat **161**: 708-734.

Hill, W.G., Goddard, M.E., Visscher, P.M. (2008). “Data and theory point to mainly additive genetic variance for complex traits.” Plos Genetics **4**: e1000008

Ihle, S., Ravaoarimanana, I., Thomas, M., Tautz, D. (2006). “An analysis of signatures of selective sweeps in natural populations of the house mouse. Mol Biol Evol **23**: 790-797.

Johnson, T., Barton, N. (2005). „Theoretical models of selection and mutation on quantitative traits.” Phil. Trans. R. Soc. B **360**: 1411–1425

Klingenberg, C.P. (2008): MorphoJ. Faculty of Life Sciences, University of Manchester, UK. http://www.flywings.org.uk/MorphoJ_page.htm

Klingenberg, C.P., Leamy, L.J., Cheverud, J.M. (2004). "Integration and modularity of quantitative trait locus effects on geometric shape in the mouse mandible." Genetics **166**(4): 1909-21.

Klingenberg, C.P., Leamy, L.J. (2001). "Quantitative genetics of geometric shape in the mouse mandible." Evolution **55**(11): 2342-2352.

Klingenberg, C.P., Leamy, L.J., Routman, E.J., Cheverud, J.M. (2001). "Genetic architecture of mandible shape in mice: effects of quantitative trait loci analyzed by geometric morphometrics." Genetics **157**: 785-802.

Klingenberg, C.P., Mebus, K., Auffray, J.C. (2003). "Developmental integration in a complex morphological structure: how distinct are the modules in the mouse mandible?" Evolution & Development **5**: 522-31.

Klingenberg, C.P., Monteiro, L.M. (2005). "Distances and directions in multidimensional shape spaces: implications for morphometric applications." Syst Biol **54**(4): 678-688.

Klingenberg, C.P. (2009). "Morphometric integration and modularity in configurations of landmarks: tools for evaluating a priori hypotheses." Evol Dev **11**: 405-421.

Lacy, R. C., Horner, B. E. (1996). "Effects of inbreeding on skeletal development of *Rattus villosissimus*." J Heredity **87**: 277-287.

Laib, A., Barou, O., Vico. L., Lafage-Proust, M.H., Alexandre, C., Rügsegger, P. (2000). "3D micro-computed tomography of trabecular and cortical bone architecture with application to a rat model of immobilisation osteoporosis." Med Biol Eng Comput **38**: 326-332.

Lande, R. (1979). „Quantitative genetic analysis of multivariate evolution, applied to brain:body size allometry." Evolution **33**: 402-416.

Lande, R., Arnold, S.J. (1983). "The measurement of selection on correlated characters." Evolution **37**: 1210-1226.

Leamy, L.J., Routman, E.J., Cheverud, J.M. (1997). "A search for quantitative trait loci affecting asymmetry of mandibular characters in mice." Evolution **51**(3): 957-969.

Leamy, L.J., Workman, M.S., Rotman, E.J., Cheverud, J.M. (2005). "An epistatic genetic basis for fluctuating asymmetry of tooth size and shape in mice." Heredity **94**: 316-325.

Leamy, L.J., Klingenberg, C.P., Sherratt, E., Wolf, J.B., Cheverud, J.M. (2008). "A search for quantitative trait loci exhibiting imprinting effects on mouse mandible size and shape." Heredity **101**(3): 518-526.

Le Roux, V., Chapuis, J.-L., Frenot, Y., Vernon, P. (2002). "Diet of the house mouse (Mus musculus) on Gouillou island, Kerguelen archipelago, subantarctic." Polar Biol **25**: 49-57.

Levrini, L., Deli, R., Sfondrini, F., Pantanali, F. (2003). "Consistency of diet and its effects on mandibular morphogenesis in the young rat." Prog Orthod **4**: 3-7.

Mao, J.J., Nah, H.-D. (2004). „Growth and development: hereditary and mechanical modulations.“ Am J Orth Dent Orth **125**: 676-689.

Macholan, M. (1996a). "Key to European house mice." Folia Zoologica **45**(3): 209-217.

Macholan, M. (1996b). "Morphometric analysis of European house mice." Acta Theriologica **41**(3): 255-275.

Macholan, M. (2008). "The mouse skull as a source of morphometric data for phylogeny inference." Zool Anz **247**: 315-327.

Macholan, M., Mikula, O., Vohralik, V. (2008). "Geographic variation of two eastern-Mediterranean non-commensal mouse species, Mus macedonicus and Mus cypriacus (Rodentia: Muridae) based on traditional and geometric approaches to morphometrics." Zool Anz **247**: 76-80.

Maynard Smith, J., Burian, P. Kauffman, S., Alberch, P., Campbell, J., Goodwin, B., Lande, R., Raup, D., Wolpert, L. (1985). “Developmental constraints and evolution.” Q Rev Biol **60**: 265-287.

McPhee, M.E. (2004). "Morphological change in wild and captive oldfield mice *Peromyscus polionotus subgriseus*." J Mammal **85**: 1130-1137.

Mitteroecker, P., Gunz, P.(2009). "Advances in Geometric Morphometrics." Evol Biol **36**: 235-247.

Maki, K., Nishioka, T., Shiori, E., Takahashi, T., Kimura, M. (2002). "Effects of diet consistency on the mandible of rats at the growth stage: computed y-ray densitometric and cephalometric analysis."

Moore, J.H., Williams, S.M. (2009). “Epistasis and its implications for personal genetics.” Am J Human Gen **85**: 309-320

Pergams, O.R.W., Ashley, M.V. (2001). "Microevolution in island rodents." Genetica **112-113**: 245-256.

Phillips, P.C. (2008). “Epistasis - the essential role of gene interactions in the structure and evolution of genetic architecture.” Nat Rev Gen **9**, 855-867.

Pialek, J., Vyskocilova, M., Bimova, B., Havelkova, D., Pialkova, J., Dufkova, P., Bencova, V., Dureje, L., Albrecht, T., Hauffe H. C., Macholan, M., Munclinger, P., Storchova, R., Zajicova, A., Holan, V., Gregorova, S., Forejt, J. (2008). "Development of unique house mouse resources suitable for evolutionary studies of speciation." J Hered **99**: 34-44.

Pigliucci, M, (2006). “Genetic variance-covariance matrices: a critique of the evolutionary quantitative genetics research program.” Biol Philos **21**: 1-23.

R Development Core Team (2008). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org.

Renaud, S., Auffray, J.-C., Michaux, J. (2006). „Conserved phenotypic variation patterns, evolution along lines of least resistance, and departure due to selection in fossil rodents." Evolution **60**: 1701-1717.

Renaud, S., Auffray, J.-C., de la Porte, S. (2010). "Epigenetic effects on the mouse mandible: common features and discrepancies in remodeling due to muscular dystrophy and response to food consistency." BMC Evolutionary Biology **10:**28

Renaud, S., Auffray, J.-C.(2009). "Adaptation and plasticity in insular evolution of the house mouse mandible." J Zool Evol Res **48**, 138-150.

Rohlf, F.J. (2005): tpsDig, digitize landmarks and outlines, version 2.05. Department of Ecology and Evolution, State University of New York at Stony Brook.

Rohlf, F.J. (2004): tpsUtil, file utility program. version 1.26. Department of Ecology and Evolution, State University of New York at Stony Brook.

Salazar-Ciudad, I. (2006). „Developmental constraints vs. variational properties: how pattern formation can help to understand evolution and development." J Exp Zool (Mol Dev Evol) **306B**: 107-125.

Salazar-Ciudad, I. (2007). "On the origins of morphological variation, canalization, robustness and evolvability." Int Comp Biol **47**: 390-400.

Schluter, D. (1996). „Adaptive radiation along genetic lines of least resistance." Evolution **50**: 1766-1774.

Scriven, P.N., Bauchau, V. (1992). "The effect of hybridization on mandible morphology in an island population of the house mouse." J Zool Lond **226**: 573-583.

Slabova, M., Frynta, D. (2007). "Morphometric variation in nearly unstudied populations of the most studied mammal: the non-commensal house mouse (Mus musculus domesticus) in the Near East and Northern Africa." Zool Anz **247**: 91-101.

Stayton, C.T. (2005). „Morphological evolution of the lizard skull: a geometric morphometrics survey." J Morph **263**: 47-59.

Tamura, K., Dudley, J., Nei, M., Kumar, S. (2007). *MEGA4*: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. Molecular Biology and Evolution **24**:1596-1599.

Tucker, P.K., Sandstedt, S.A., Lundrigan, B.L. (2005). "Phylogenetic relationships in the subgenus Mus (genus Mus, family Muridae, subfamily Murinae): examining gene trees and species trees." Biol J Linn Soc **84**: 653-662.

Turelli, M., Barton, N.H. (2006). „Will population bottlenecks and multilocus epistasis increase additive genetic variance?" Evolution **60**: 1763-1776.

Waddington, C.H. (1942). "Canalization of development and the inheritance of acquired characters." Nature **150**: 563-565.

Wagner, G.P., Mezey, J. (2000). “Modelling the evolution of genetic architecture: a continuum of alleles model with pairwise A x A epistasis.” J Theor Biol **203**: 163-175.

Wroe, S., Milne, N. (2007). “Convergence and remarkably consistent constraint in the evolution of carnivore skull shape.” Evolution **61**: 1251-1260.

Yamada, K., Kimmel, D.B. (1991). "The effect of dietary consistency on bone mass and turnover in the growing rat mandible." Arch Oral Biol **36**: 129-138.

Zelditch, M., Swiderski, D., Sheets, D.H., Fink, W. (2004) Geometric morphometrics for biologists. Elsevier Academic Press, New York, London.