8-24-2020

Inferred Genetic Architecture Underlying Evolution in a Fossil Stickleback Lineage

Yoel E. Stuart
Loyola University Chicago, ystuart@luc.edu

Matthew P. Travis
Rowan University

Michael A. Bell
University of California Museum of Paleontology

Recommended Citation

Stuart, Yoel E.; Travis, Matthew P.; and Bell, Michael A.. Inferred Genetic Architecture Underlying Evolution in a Fossil Stickleback Lineage. Nature Ecology and Evolution, 4, : pages1549–1557, 2020. Retrieved from Loyola eCommons, Biology: Faculty Publications and Other Works, http://dx.doi.org/10.1038/s41559-020-01287-x

Title: Inferred genetic architecture underlying evolution in a fossil stickleback lineage

Authors: Yoel E. Stuart[1,*], Matthew P. Travis[2], and Michael A. Bell[3]

[1] Department of Biology, Loyola University Chicago, Chicago, IL, USA

[2] Department of Biological Sciences, Rowan University, Glassboro, NJ, USA

[3] University of California Museum of Paleontology, Berkeley, CA, USA

[*] Corresponding author: ystuart@luc.edu ORCID ID: 0000-0003-1235-6277

Inferring the genetic architecture of evolution in the fossil record is difficult because genetic crosses are impossible, the acquisition of DNA is usually impossible, and phenotype-genotype maps are rarely obvious. However, such inference is valuable because it reveals the genetic basis of microevolutionary change across many more generations than is possible in studies of extant taxa, thereby integrating microevolutionary process and macroevolutionary pattern. Here, we infer the genetic basis of pelvic skeleton reduction in *Gasterosteus doryssus*, a Miocene stickleback fish from a finely resolved stratigraphic sequence that spans nearly 17,000 years. Reduction in pelvic score, a categorical measure of pelvic structure, resulted primarily from reciprocal frequency changes of two discrete phenotypic classes. Pelvic vestiges also showed left-side-larger asymmetry. These patterns implicate *Pitx1*, a large-effect gene whose deletion generates left-larger asymmetry of pelvic vestiges in extant, closely-related *Gasterosteus aculeatus*. In contrast, reductions in lengths of the pelvic girdle and pelvic spines resulted from directional shifts of unimodal, continuous trait distributions, suggesting an additional suite of genes with minor, additive pelvic effects, again like *G. aculeatus*. Similar genetic architectures explain shared but phyletically independent patterns across 10 million years of stickleback evolution.

We studied the last ~16,500 years of a ~108,275 year-long fossil *G. doryssus* sequence[1,2]. The entire sequence contains two lineages of *G. doryssus*. Lineage I existed during the first 92,012 years of the sequence and had a vestigial pelvic girdle and fewer than three dorsal spines, on average. At 92,012 years, lineage I was replaced within 125 years by lineage II (Extended Data 2), whose source was a parapatric *G. doryssus* population from outside the depositional basin[3]. At the replacement event, lineage II invariably had a robust pelvis and three dorsal spines[2].

Lineage II subsequently evolved vestigial armor phenotypes, similar to those of lineage I[1], under directional natural selection[4] over 16,750 years[3]. Armor reduction in Lineage II included reduction in the size and complexity of the pelvic girdle and pelvic spines.

Such pelvic reduction exists in many extant lake populations of *G. aculeatus*[5] and is likely also driven by natural selection[6,7]. The major gene underlying pelvic reduction in *G. aculeatus* is usually *Pitx1* (*Pituitary homeobox 1*)[7-12] (but see [13]). Loss occurs through deletion mutations in the *Pel* enhancer region that reduce pelvis-specific *Pitx1* expression[7,12]. Deletion mutations of *Pel* act recessively and their phenotypic effects on pelvic score (PS: a categorical metric of pelvis size and complexity; Methods) segregate in multimodal, near Mendelian fashion in *G. aculeatus*[9,10]. Moreover, because a paralogous gene, *Pitx2*, presumably also contributes to pelvic girdle formation but is expressed more on the left side than the right[14], reduced *Pitx1* expression results in a directionally asymmetrical, left-larger pelvic vestige in *G. aculeatus*[5,9,11,13,15] (as well as other vertebrates[9,14,16]). Because *Pitx1* has repeatedly played a major role in pelvic reduction in *G. aculeatus*, and because *G. aculeatus* is closely related to *G. doryssus*[17], *Pitx1* is a good candidate gene for pelvic reduction in *G. doryssus*. To infer whether *Pitx1* is responsible for major pelvic reduction in *G. doryssus*[1,3], we examined our fossil data for a Mendelian pattern of pelvic scores and for left-larger pelvic vestige asymmetry (**Question 1**).

Pelvic reduction in extant *G. aculeatus* also mapped to several genomic regions with minor, additive effects[9-11,18]. To infer whether genes with small effects also contributed to pelvic reduction in *G. doryssus* (**Question 2**), we used the fact that pelvic score in *G. doryssus* did not decline immediately after the fully armored lineage II appeared in the temporal sequence[3]. Although reduction of other, non-pelvic armor traits (*i.e.*, numbers of dorsal spines and touching predorsal pterygiophores) began to evolve immediately[3], pelvic score remained static in lineage II for ~3,500 years before declining. We examined new data for pelvic girdle length and pelvic spine length to ask whether those traits had also experienced delayed reduction. If yes, this suggests that natural selection for armor loss did not initially include selection for reduced pelvises. If not, however, and reduction began immediately, this suggests that pelvic reduction in *G. doryssus* was polygenic and that there was variation in minor genes that allowed some pelvic reduction in response to natural selection before the appropriate, hypothesized mutations in *Pitx1* arose and became the basis for more extensive pelvic reduction.

Results

Question 1. Is *Pitx1* responsible for the reduction of pelvic score observed in lineage II of *G. doryssus*, ~3500 years after lineage II appeared? We measured the lengths of the right and left pelvic vestiges from 815 specimens from temporal sequence L (Methods, Extended Data 2). A paired t-test indicated that pelvic vestiges were significantly larger on the left side (mean right-minus-left difference = -0.29mm, $t_{814} = -8.39$, $p < 0.0001$). The left vestige was larger in 73.5% of pelvic-reduced specimens, significantly more than half ($\chi_1^2 = 95.31$, $p < 0.0001$; Figure 1), corroborating a preliminary finding using a much smaller sample size[13]. This left-bias was not strongly influenced by the extent of pelvic reduction (*i.e.*, pelvic score category; Methods; $\chi_9^2 = 13.05$, $p = 0.16$, Table S1).

The distributions of pelvic scores (PS) of lineage II specimens were multimodal during reduction. In temporal sequence L, all but three of the 595 specimens in the first 12 samples following replacement of lineage I by lineage II had a full pelvis (PS 3.0; Figure 2), spanning the first 2,750 years of lineage II (Table S2). Then, mean PS declined to PS 2.92, where it remained static for another 750 years (Figure 2, Table S2). This decline to 2.92 was caused by appearance of only two specimens with extreme pelvic reduction (PS $\leq$ 1) out of 66 (Table S2). After another 500 years, a third mode formed at PS 2.0, driven mostly by an increase of fish with vestigial pelvises of PS 1.0 (Figure 2, Table S2). PS declined after that to a mean value of about 1.0 by 10,000 years after replacement (Figure 2, Table S2). This new phenotype is indistinguishable from pelvic vestiges that had characterized lineage I for 92,012 years before lineage II appeared[1,3]. Specimens that lack the pelvis entirely appear near the middle of the sequence but never become very frequent (maximum PS 0 = 16.7%, 13,750 years after replacement; Table S2). We are unsure why PS 3.0 individuals do not disappear completely, though low frequency dispersal from the lineage II source population could explain this pattern; occasional full-pelvis migrants were also detected in the first 92,012 years of the fossil sequence[1] (Extended Data 2; Methods).

Our evidence thus suggests that *Pitx1* was indeed the major gene responsible for pelvic reduction in *G. doryssus* lineage II. First, the reduction in mean pelvic score through time (Figure 2, Figure 4) resulted largely from changes in the relative frequencies of specimens in two discrete, contrasting phenotypic classes (i.e., PS 1.0 and 3.0; Figure 3, Table S2) rather than from a gradual change from PS 3.0 toward PS 1.0 in the position of a single mode. This is

consistent with the action of an allele of large effect, like deletion of a pelvic enhancer of *Pitx1*[7,9,15]. Second, left-larger directional asymmetry of pelvic vestiges (Figure 1) implicates *Pitx1* as the primary factor for pelvic reduction in lineage II of *G. doryssus* because (i) only six genes in the vertebrate genome are known to be related to directional asymmetry in limb bud tissue during development[14]; (ii) left-biased asymmetry is a known outcome of deletion mutations that reduce expression of *Pitx1* in vertebrates[15]; and (iii) in extant *G. aculeatus*, vestigial pelvic phenotypes that map to the *Pitx1* locus tend to be larger on the left side[9].

Third, the lag time of 3500 years observed before reduction of pelvic score in *G. doryssus* lineage II (Figure 2, Figure 4) can be explained by a recessive mutation like *Pitx1*, as follows. Natural selection favored armor reduction in lineage II immediately following its replacement of lineage I[3,4 5,6,19 20,21]. This is evident from immediate reduction in the mean number of dorsal spines (armor structures[22]), the mean number of touching pre-dorsal pterygiophores[3] (which are structurally and likely functionally related to the dorsal spines), as well as in mean pelvic girdle length and mean pelvic spine length (this study; see Question 2, Figure 4). Despite selection for armor reduction[4], however, pelvic score remained at 3.0 for several thousand years (Table S2). This suggests that the founders of lineage II initially lacked a *Pitx1* allele for pelvic reduction, or carried it at such low frequencies that individuals that expressed the reduced allele (ie., homozygotes) were too rare for directional selection to act efficiently. This could make sense because the parapatric source population for lineage II could have been under purifying selection to remove pelvic-reducing deletion mutations of large effect in *Pitx1*[2]; this source population coexisted with predatory fishes that were present elsewhere in the larger drainage (but not in the depositional environment sampled here) [2,23,24]. Fish predators select for armor [5,6,19-21].

Eventually, however, reduction of pelvic score proceeded in Lineage II based on alleles that reduce *Pitx1* expression during development of the pelvis. Where would low-armor *Pitx1* variants have come from? In extant *G. aculeatus*, *de novo* deletion mutations of the *PelA* enhancer occur at a remarkably high rate[12]. *PelA* lies within a stretch of fragile DNA that experiences deletion mutations nearly four orders of magnitude faster than in other parts of the threespine stickleback genome and than is typical of vertebrate genomes[12], increasing the likelihood that enhancer mutations will be generated. Several *G. aculeatus* populations have evolved pelvic loss over the last 15,000 years by independently acquiring deletion mutations in the *PelA* enhancer region of *Pitx1*[7]. This suggests that appropriate *Pitx1* mutations in *G. doryssus*

lineage I could have arisen often in the environment sampled here. However, if they were also recessive (as in modern stickleback[9,10]), these mutations would have had to drift to an appreciable frequency before homozygotes would occur and selection could drive the deletion mutation toward fixation. This drift component of the fixation process is consistent with the occasional appearance of vestigial pelvises during early lineage II samples when mean pelvic score remained near 3 (Table S2, Figure 2). A delay of ~3500 years before pelvic reduction is within the range of simulated lag times in pelvic reduction found by population genetic modeling by Xie et al. [12] (see figure 4D in [12] and figures S6 and S7 in the Supplementary Materials of [12]), given known mutation rates in the *Pel* enhancer region, reasonable selection coefficients for pelvic loss ($0.1 > s > 0.01$), and relevant *G. aculeatus* population sizes. They found that the probability of generating and fixing a *de novo* mutation in a fragile genome region was 1.0 for reasonable stickleback population sizes ($10^3 < N < 10^6$) within 10,000 generations. When selection was $s = 0.01$, predicted time to fixation was less than 5000 generations. When selection was $s = 0.1$, predicted time to fixation was less than 2000 generations. Assuming that *G. doryssus* had a generation time of two years[25], the observed delay of ~3500 years to begin reduction and then another ~3000 years to reach a mode of PS 1 is reasonably close to population genetic modeling for *de novo* mutation in the *Pel* enhancer region.

Thus, we conclude that major pelvic reduction in lineage II *G. doryssus* likely depended on a new mutant (or very rare standing) recessive allele of *Pitx1*.

Question 2. Does immediate reduction in pelvic girdle length and pelvic spine length reveal the presence of alleles of minor, additive effects? We used samples from a different temporal sequence, K (Methods), but the same section of rock to answer this question. We first tested whether pelvic girdle length and pelvic spine lengths declined immediately following the replacement event of lineage I by lineage II (i.e., immediately have a negative slope for the trait mean vs. time), while pelvic score delayed reduction (i.e., has an initial slope of zero) . To do this, we used a piecewise ("broken-stick") regression model, which estimates the slope and intercept for two different pieces of a regression line, before and after a break point where the slope is estimated to change significantly. Consistent with the visually obvious lag time before pelvic score reduction (Figure 4), the first 'stick' inferred by the piecewise regression for pelvic score against time had an intercept of 3.0 (the maximum possible pelvic score), a slope of zero,

and a breakpoint between temporal samples five and six (Table 1). After this breakpoint, the slope coefficient became negative (Table 1). In contrast, for fish with PS 3.0 (*i.e.*, when the fully functional dominant allele of *Pitx1* is hypothesized to be at high frequency), mean size-corrected lengths of both pelvic girdle and pelvic spines began to decline immediately after appearance of lineage II (Figure 4). The slope of the first 'stick' was significantly negative for both traits (Table 1). This implies available genetic variation unlinked to *Pitx1*.

Moreover, these significant trends for reduction of mean size-corrected pelvic girdle and pelvic spine lengths both resulted from gradual shifts to smaller sizes by unimodal (Figure 5, Table S3), normally distributed frequency distributions (Table S3). This finding is consistent with multiple genes acting additively. Pearson correlations between pelvic girdle and pelvic spine lengths calculated for each of the first ten samples in temporal sequence K averaged only 0.38 (sd = 0.28; max = 0.74; min = -0.19), suggesting that the two traits might be reduced in part via different genetic changes. A QTL study in *G. aculeatus* from a cross between populations with complete and missing pelvises found that the two traits shared four QTL for length, but that pelvic spine length also has a unique QTL that explains 5.6% of its variance[9]. Thus, in that QTL cross at least, there was potential for independent variation in the lengths of pelvic spine and pelvic girdle, consistent with observation in the fossils.

We further measured pelvic vestige lengths for a subsample of 305 fossils with PS 1.0 (Table S4)—that is, individuals likely to have been homozygous for a null allele of *Pitx1* in the pelvis. The distribution of lengths did not deviate from unimodal (Dip statistic $D_n = 0.02$, p = 0.67). Last, we note that the pelvis did not completely disappear once the hypothesized deletion mutations arose in *Pitx1*; *i.e.,* PS 0 was not common. The persistence of intermediate pelvic scores (i.e., PS 2.8 to 1.2; Figure 2, Figure 3) and the unimodal distribution of vestigial pelvic girdle lengths in fish with PS = 1.0 further suggest that other genes besides *Pitx1* were also involved in pelvic development and reduction in *G. doryssus*.

Thus, we infer that *Pitx1* likely was not the sole genetic cause for pelvic reduction in lineage II. Pelvic reduction also involved a suite of additive alleles with small effects. Such alleles in lineage II *G. doryssus* would rarely produce strong pelvic reduction in any one individual and could be carried even when selection favored full pelvises, as in the putative source population of lineage II[2]. However, once selection for pelvic reduction began in lineage II of *G. doryssus* (*i.e.*, following appearance of lineage II to our depositional environment), these

187 loci would have facilitated immediate reduction of pelvic girdle and pelvic spine lengths (Figure
188 4).
189 This inference is consistent with evolution in *G. aculeatus*, in which quantitative trait loci
190 (QTL) with small, additive effects on pelvic elements contribute to pelvic reduction[9-11,18].
191 However, we note that there are phenotypic differences in the order of structural reduction of the
192 pelvic skeleton between extant *G. aculeatus* and fossil *G. doryssus*[26]. In *G. doryssus,* the pelvic
193 spine is lost first, at which point the pelvic girdle breaks into separate anterior and posterior
194 elements that correspond to different developmental structures[27]. The size of the vestigial
195 posterior element can vary in the fossils, but it is usually absent. In fossil specimens with PS 1.0
196 (*i.e.*, no posterior element), the anterior element varies in size and can also be lost unilaterally or
197 on both sides. In contrast, in extant *G. aculeatus*, pelvic reduction usually proceeds through loss
198 of the pelvic spine without the pelvic girdle dividing into separate anterior and posterior
199 elements. Following spine loss, the posterior process gets shorter, leaving only a diminutive
200 ascending branch emanating from the anterior process. Next, the ascending branch gets shorter,
201 until it eventually leaves a structure that is indistinguishable from the tear-drop shaped anterior
202 element in the fossils. Finally, like the fossils, the anterior element is reduced in size and lost
203 unilaterally or bilaterally. These phenotypic differences in the order of loss and the separation of
204 anterior and posterior pelvic elements suggests that the number, identity, and expression of small
205 effect genes differs between *G. aculeatus* and *G. doryssus*. However, in both species, it is the
206 posterior half of the pelvic girdle that is most often reduced or missing. In *G. aculeatus,* the
207 posterior process develops separately from the anterior process [27] and it is thus likely that the
208 posterior and anterior processes in *G. doryssus* also are underlain by separate developmental
209 modules.
210
211 **Discussion**
212 Despite being separated by 10 million years, our data suggest that *G. aculeatus* and *G. doryssus*
213 have both used *Pitx1* during evolution of major reduction of their pelvic armor. Inference of the
214 gene(s) responsible for skeletal change in the fossil record is very rare. For example, Schmid and
215 Villagra[28] attributed discontinuities in scale and skeletal variation among species of Triasssic
216 *Saurichthys* to two growth factors (i.e., *Ectodysplasin*, *Fibroblast Growth Factor*) or their
217 receptors. They argued that involvement of these genes in development of homologous structures

in extant species implicates them in evolution of *Saurichthys* morphology. However, the temporal, phylogenetic, and morphological differences between *Saurichthys* and the modern analogues allows only preliminary conclusions. Similarly, Meredith et al.[29] document repeated transitions in the gene *enamelin* from a functional gene in extant mammals with enameled teeth to a pseudogene in those lacking enamel or without teeth. They suggested *enamelin* was likely to have been responsible for losses in the fossil record. Qu et al. [30] made a similar argument for the role of *enamalin* and several other genes during tooth gain and loss in stem osteichthyans. Finally, Zhu and colleagues [31,32] proposed that loss of *sparc1* in stem Chondrichthyans caused a secondary loss of perichondral bone in that clade. We were not able to find additional, relevant examples during a literature search in March 2019, searching "fossil gene*" and related queries on scholar.google.com. (We did find, however, that that inferences of broader genetic architecture responsible for change in the fossil record are more common (*e.g.*[33-37]).)

This paucity of examples arises in part because claiming that a specific gene caused phenotypic variation in a fossil lineage assumes that no other regions of the genome can generate similar phenotypic effects. In other words, a plausible but ultimately untestable alternative hypothesis exists: a different gene(s) was involved. Indeed, for stickleback, evidence is accumulating that phenotypic parallelism does not necessarily imply genetic parallelism[38,39]. However, for the following four reasons, we argue that our evidence meets a reasonable burden of proof to infer the role of a specific gene, *Pitx1*, in pelvic reduction in a fossil species. (i) First, many genes involved in pelvis development also play a role in development elsewhere in the body. *Pitx1* is no exception and is expressed in the jaw, pituitary gland, and other tissues during development; mice with null mutations in the coding region of *Pitx1* die before birth or as neonates and exhibit developmental abnormalities of the jaw, pituitary, and other structures[7,40,41]. However, *Pitx1* stands out among candidates because its expression can be modulated specifically in the pelvis without disrupting development elsewhere. Mutations to the *PelA* enhancer region reduce expression of *Pitx1* in the pelvis[7,12]. (ii) Second, reduction of *Pitx1* expression is clearly involved in generating left-larger asymmetry in hindlimb elements through *Pitx1*'s interaction with *Pitx2*[15]. *Pitx2* is one of only six genes known to generate left-larger directional asymmetry in vertebrate lateral plate mesoderm, the source of limb buds[14]. (iii) Third, the *PelA* enhancer of *Pitx1* lies in a fragile portion of the genome that shows mutation rates ~4 orders of magnitude higher than background[12] and shows signatures of positive natural

selection[7,12], suggesting that favored variation might arise often at this locus. (iv) Fourth, pelvic loss in Canadian and European populations of ninespine stickleback (*Pungitius pungitius*) maps to *Pitx1,* suggesting that a parallel genetic mechanism for pelvic loss persisted across at least 7.2 to 6.9 million years of divergence from a common ancestor with *G. aculeatus*[42,43]; this timescale is similar to our comparison between fossil *G. doryssus* and *G. aculeatus*. This result, combined with the repeated use of *Pitx1* during pelvic loss by multiple independent populations of *G. aculeatus* as well as in manatees[9,11,16] suggests that *Pitx1*'s role in pelvic reduction can be remarkably parallel across distantly related and phenotypically diverse vertebrates. Thus, though we can never disprove the alternative hypothesis that a different gene causes parallel phenotypic outcomes in fossil *G. doryssus* and extant *G. aculeatus*, we feel that such a hypothesis is less plausible than the simpler conclusion: *Pitx1* is the likely gene of major effect in this fossil system.

Methods

The Fossil System

The fossil stickleback *Gasterosteus doryssus* (Extended Data 1) is abundant and well preserved in a Miocene (10 million year old) lake deposit with annual layers, providing both excellent samples and fine temporal resolution (reviewed by [2]). We focused on the evolution of lineage II because we could observe evolution from an armored form, with full pelvic girdles and both pelvic spines, to a vestigial form with reduced pelvic girdles and fewer, smaller pelvic spines.

Location and fossil sampling

Fossil *G. doryssus* were collected from an open pit, diatomaceous earth mine at 39.526° N, 119.094° W, near Reno, Nevada, USA. In the field, we used sharpened putty knives to split the rock along arbitrary bedding planes to find fossils. Each fossil's approximate stratigraphic position was measured in relation to volcanic ash layers. Specimens were prepared in the laboratory under a dissecting microscope, using probes to remove the matrix that covered bones[1,3]. All specimens of *G. doryssus*, as well as lithological samples and associated field notes, have been deposited in University of California Museum of Paleontology.

Temporal Sequence Correlations

Note that a 'section' is a span of stratigraphic thickness of rock. A 'sample' comprises multiple fossil specimens that are all mined from the same section. Multiple samples make up a 'temporal sequence'.

Fossil stickleback specimens used in this study come from two temporal sequences, K[2] and L[3], which comprise separate specimens collected with different sampling designs. However, K and L came from the same stratigraphic section in the same exposure, they overlap in time, and they occupy the upper 17% of the stratigraphic section covering temporal sequence D, reported by [1]. D includes 26 samples made mostly at 5000-year intervals and spans an estimated 108,275 years (Extended Data 2). Temporal sequence K spans 16,363 years (Extended Data 2) and comprises 18 samples made at about 1000 year intervals (Table S4, Table S5). Each sample was made from a narrow time interval of one to several consecutive years. L is one continuous sequence spanning about 21,250 years (Extended Data 2). Following Bell et al. [3], we binned specimens from L into 250-year samples for analysis (Table S2). D, L, and K can be correlated ($\pm$75 years) by aligning replacement of lineage I by lineage II observed in all three sequences. This replacement event occurs ~92,012 years after the start of D[1] (Extended Data 2).

Data use

We used existing pelvic score data [3] and new left-right pelvic vestige length data from lineage II fossils from temporal sequence L to characterize the presence of pelvic-score multimodality and directional asymmetry of pelvic vestiges (Question 1). We used existing and new data from lineage II fossils from temporal sequence K to test whether the lengths of the pelvic girdle and the pelvic spines began to decline immediately after lineage II replaced lineage I, and to infer whether the evolution of these traits is consistent with polygenic, additive genetic architecture (Question 2).

Phenotyping

Ordinal pelvic scores (PS) were assigned by MAB to pelvic phenotypes by visual inspection of all fossils in both L and K, using marked figures from reference [26] as a standard (Extended Data 3). An individual with a full pelvic girdle (*i.e.*, anterior and posterior processes, ascending branch) and both pelvic spines present was scored PS 3.0. Reduction from 3.0 always starts with loss of pelvic spines and concurrent division of the pelvic girdle into anterior and posterior

311 elements on at least one side. Pelvic scores from 2.8 to 1.2 were assigned in intervals of 0.2
312 points based on the size and complexity of the posterior process of the pelvic girdle. Reduction
313 of PS from 1.2 to 1.0 indicates that the anterior pelvic plate vestige is present on at least one side
314 but posterior vestiges have been completely lost. The jump from PS 1.0 to PS 0 indicates loss of
315 anterior pelvic plate elements on both sides[3]. Extended Data 3 provides drawings from reference
316 [26] and photograph examples to illustrate how PS was scored. PS is significantly correlated with
317 digitized pelvic girdle area between PS 1.0 to 3.0 ($r^2 = 0.82$)[3]. Between PS 2.8 and 1.0, PS
318 mostly reflects a continuous distribution of size of the posterior process of the pelvic vestige. PS
319 compresses the phenotypic scale between scores of 0 and 1.0 because it does not take into
320 account whether one or both sides of the anterior pelvic vestige are present or the size of the
321 vestige within this range.
322 For specimens in temporal sequence K, standard length was measured as the distance
323 from the tip of the upper jaw to the end of the last vertebra (hypural plate), using 'measure mode'
324 in tpsDIG[44] on digital images of each fossil. Specimens with gaps between the vertebrae were
325 excluded, and protrusion of the premaxilla was taken into account. Standard length was often
326 measured in segments to limit the effect of postmortem (*i.e.,* taphonomic) curvature of the
327 vertebral column. Pelvic girdle and pelvic spine lengths were also measured using tpsDIG.
328 Pelvic girdle lengths were measured differently depending on PS. Specimens with a full pelvis
329 (PS 3.0) were measured along the midline from the most anterior point of the anterior process to
330 the posterior tip of the posterior process of the pelvis on the side with the best preservation[45]. The
331 pelvic vestige of specimens with PS 1.0 was measured from the pointed anterior tip to the most
332 distal point on the rounded posterior edge. In specimens with PS 1.2 to 2.8, the length of the
333 posterior element along the median edge was measured and added to the length of the anterior
334 element. Specimens with no pelvic vestige (PS 0) were assigned a value of 0.0. Pelvic spine
335 lengths were measured from distal tip to the proximal base of the condyles by which they
336 articulate with the pelvic girdle.
337 For a subset of specimens in temporal sequence L, the lengths of the anterior and
338 posterior pelvic vestiges were measured as described above for K specimens. However, unlike
339 for K, we measured both the right and left sides for specimens in which overlap of the vestiges
340 between sides allowed us to distinguish right from left.
341

Analysis

All analysis was conducted in R version 3.6.1, (2019)[46]. Unless noted otherwise, statistical functions come from this version's base 'stats' package. Functions are indicated by *italics.* The statistical analysis described here was not preregistered.

Question 1. Is *Pitx1* responsible for the major pelvic reduction observed in lineage II of *G. doryssus*? We estimated directional asymmetry from the pelvic vestiges of temporal sequence L specimens for which pelvic vestige lengths were measured on right and left sides. We excluded fish with pelvic scores of 3.0, as they would not have had the hypothesized deletion mutation in *Pel* that reduces *Pitx1* expression[7,12] and analyzed only fish with pelvic scores less than 3.0 and greater than or equal to 1.0. We summed the lengths of vestigial anterior and posterior pelvic elements on the same side before quantifying length asymmetry between sides. For fish with pelvic scores of 1 (*i.e.*, no posterior elements), we compared length asymmetry in anterior elements only.

Directional asymmetry was calculated as percent asymmetry,

$$[rpv - lpv] / [rpv + lpv] * 100,$$

where rpv and lpv are the right and left pelvic vestige lengths, respectively. Thus, specimens with a larger left vestige had negative asymmetry values. We used a two-sided paired t-test (*t.test*) to test whether right versus left pelvic vestige lengths are significantly asymmetric. We used a two-sided Chi-square test (*chisq.test*) to test whether the number of specimens with larger right or left vestiges deviated significantly from 50%.

We also used a two-sided Chi-square to test whether the frequencies of specimens with larger and smaller left vestiges were influenced by pelvic score. That is, we asked if frequency distributions of vestiges with a larger pelvic vestige on the left or right side within each pelvic score class (*i.e.*, 1.0, 1.2, … 2.6, 2.8) differed from the pooled frequency distribution (Table S1). If all vestigial pelvic phenotypes (i.e., PS 1.0 to 2.8) are caused by reduction of *Pitx1* expression during pelvic girdle development, we would not expect the frequency distribution to vary among pelvic score categories[9].

Finally, if recessive alleles of a gene of large effect (*i.e., Pitx1*) underlie pelvic score evolution, then we would expect frequency distributions of pelvic score to have discrete peaks, deviating from unimodality. We verbally described the reduction in mean pelvic score through

373 time using the fine-scale temporal resolution of temporal sequence L. We illustrated the change
374 in frequency distributions of pelvic score through time within temporal sequence K, using its
375 chronologically discrete and more sparse sampling for clearer plotting and presentation (Figure
376 3).
377
378 Question 2. Does reduction in continuous pelvic traits implicate the action of genes with minor,
379 additive effects?
380 To examine the role of minor genes in *G. doryssus* pelvic evolution, we measured the
381 lengths of one pelvic spine and of the pelvis (as described above) in temporal sequence K
382 specimens with PS 3.0. We used just PS 3.0 individuals to infer whether genes with small effects
383 contributed to pelvic reduction before a gene with major effects on PS arose to obscure the
384 effects of the minor genes. We restricted our analysis to the first 10 samples of K, as only those
385 samples included enough specimens (i.e., 5 or more) with full pelvic scores to compute
386 reasonable means for pelvic spine and pelvic girdle lengths. Lengths for both continuous traits
387 were size-corrected using standard length, following[47-49] (Supplementary Information).
388 We plotted means for pelvic score and the two size-corrected traits against time to
389 visualize the timing of reduction of pelvic spine length, pelvic girdle length, and pelvic score
390 after lineage II appeared. For statistical support, we fit piecewise regressions (*i.e.,* "broken stick"
391 models) to the trait means. If pelvic girdle and spine lengths dropped immediately while pelvic
392 score remained static, the first 'stick' for both pelvic spine and girdle lengths would have a
393 significantly negative slope, while the first stick for pelvic score would have a slope of zero. For
394 each trait, we modelled the linear relationship between mean trait values and time since lineage
395 II first appeared. Each model allowed one breakpoint along the temporal sequence of samples,
396 such that we estimated two sets of slopes and intercepts before and after the proposed breakpoint.
397 For each trait, we iterated through models that differed by where in the temporal sequence the
398 breakpoint was proposed. Then we chose the model with the lowest residual error as our best
399 estimate for the first temporal breakpoint. We limited our potential breakpoints to the first seven
400 samples because visual inspection suggests that significant differences in 'first stick' slope
401 between pelvic score and the other traits occur in this span (Figure 4). Moreover, the eighth
402 sample contains an increase in trait means (Figure 4). We note that with only seven values (*i.e.*,

seven samples), significance tests of slope and intercept have low power. Thus, we are mainly interested in the sign and estimate of the model parameters.

Next we plotted the frequency distributions for pelvic spine length and pelvic girdle length within each of the first 10 temporal samples for only individuals with PS 3.0. If the minor alleles underlying evolution of these traits are additive, we would expect these distributions to be normal and unimodal. We used Shapiro-Wilk Normality tests (*shapiro.test*) within each sample for each trait to test for deviations from normality. Complementarily, we used Hartigan's Dip Statistic, D_n[50], to test for deviations from unimodality (*dip.test* in the package 'diptest' [51]). We calculated Pearson correlations (*cor*) to quantify the relationship between pelvic spine length and pelvic girdle length for each sample; a strong correlation might imply the same genetic mechanism for reduction. Finally, we pooled all individuals from temporal sequence K with PS 1.0 (*i.e.*, likely homozygous for the null *Pitx1* allele) and, as above, asked whether pelvic girdle length was unimodal and normal. If so, it would further corroborate evidence that minor alleles contributed to pelvic reduction.

Data and Code Availability

Data and code are available at datadryad.org (https://doi.org/10.5061/dryad.02v6wwq18).

References

1. Bell, M. A., Baumgartner, J. V. & Olson, E. C. Patterns of temporal change in single morphological characters of a Miocene stickleback fish. *Paleobiology* **11,** 258–271 (1985).
2. Bell, M. A. Implications of a fossil stickleback assemblage for Darwinian gradualism. *Journal of Fish Biology* **75,** 1977–1999 (2009).
3. Bell, M. A., Travis, M. P. & Blouw, D. M. Inferring natural selection in a fossil threespine stickleback. *Paleobiology* **32,** 562–577 (2006).
4. Hunt, G., Bell, M. A. & Travis, M. P. Evolution toward a new adaptive optimum: phenotypic evolution in a fossil stickleback lineage. *Evolution* **62,** 700–710 (2008).
5. Klepaker, T., Østbye, K. & Bell, M. A. Regressive evolution of the pelvic complex in stickleback fishes: a study of convergent evolution. *Evolutionary Ecology Research* **15,** 413–435 (2013).
6. Bell, M. A., Ortí, G., Walker, J. A. & Koenings, J. P. Evolution of pelvic reduction in threespine stickleback fish: a test of competing hypotheses. *Evolution* **47,** 906–914 (1993).
7. Chan, Y. F. *et al.* Adaptive evolution of pelvic reduction in sticklebacks by recurrent deletion of a *Pitx1* enhancer. *Science* **327,** 302–305 (2010).

8. Cole, N. J., Tanaka, M., Prescott, A. & Tickle, C. Expression of limb initiation genes and clues to the morphological diversification of threespine stickleback. *Current Biology* **13,** R951–R952 (2003).
9. Shapiro, M. D. *et al.* Genetic and developmental basis of evolutionary pelvic reduction in threespine sticklebacks. *Nature* **428,** 717 (2004).
10. Cresko, W. A. *et al.* Parallel genetic basis for repeated evolution of armor loss in Alaskan threespine stickleback populations. *Proc. Natl. Acad. Sci. U.S.A.* **101,** 6050–6055 (2004).
11. Coyle, S. M., Huntingford, F. A. & Peichel, C. L. Parallel evolution of *Pitx1* underlies pelvic reduction in Scottish Threespine Stickleback (*Gasterosteus aculeatus*). *Journal of Heredity* **98,** 581–586 (2007).
12. Xie, K. T. *et al.* DNA fragility in the parallel evolution of pelvic reduction in stickleback fish. *Science* **363,** 81–84 (2019).
13. Bell, M. A., Khalef, V. & Travis, M. P. Directional asymmetry of pelvic vestiges in threespine stickleback. *Journal of Experimental Zoology Part B: Molecular and Developmental Evolution* **308B,** 189–199 (2007).
14. Palmer, A. R. Symmetry breaking and the evolution of development. *Science* **306,** 828–833 (2004).
15. Marcil, A., Dumontier, É., Chamberland, M., Camper, S. A. & Drouin, J. Pitx1 and Pitx2 are required for development of hindlimb buds. *Development* **130,** 45–55 (2003).
16. Shapiro, M. D., Bell, M. A. & Kingsley, D. M. Parallel genetic origins of pelvic reduction in vertebrates. *Proc. Natl. Acad. Sci. U.S.A.* **103,** 13753–13758 (2006).
17. Bell, M. A. in *The Evolutionary Biology of the Threespine Stickleback* (eds. Bell, M. A. & Foster, S. A.) 438–471 (Oxford University Press, 1994).
18. Peichel, C. L. *et al.* The genetic architecture of divergence between threespine stickleback species. *Nature* **414,** 901–905 (2001).
19. Rollins, J. L., Lohman, B. K. & Bell, M. A. Does ion limitation select for pelvic reduction in threespine stickleback (*Gasterosteus aculeatus*)? *Evolutionary Ecology Research* **16,** 101–120 (2014).
20. Reimchen, T. E. Spine deficiency and polymorphism in a population of *Gasterosteus aculeatus*: an adaptation to predators? *Canadian Journal of Zoology* **58,** 1232–1244 (1980).
21. Reimchen, T. E. in *The Evolutionary Biology of the Threespine Stickleback* (eds. Bell, M. A. & Foster, S. A.) 240–276 (Oxford University Press, 1994).
22. Hoogland, R., Morris, D. & Tinbergen, N. The spines of sticklebacks (*Gasterosteus* and *Pygosteus*) as a means of defense against predators (*Perca* and *Esox*). *Behaviour* **10,** 205–236 (1956).
23. Baumgartner, J. V. A new fossil ictalurid catfish from the Miocene middle member of the Truckee Formation, Nevada. *Copeia* **1982,** 38–46 (1982).
24. Stearley, R. F. & Smith, G. R. Fishes of the Mio-Pliocene western Snake River plain and vicinity. *Miscellaneous Publications Museum of Zoology of the University of Michigan* **204,** 1–43 (2016).
25. Baker, J. A. in *The Evolutionary Biology of the Threespine Stickleback* (eds. Bell, M. A. & Foster, S. A.) 144–187 (Oxford University Press, 1994).
26. Bell, M. A. Interacting evolutionary constraints in pelvic reduction of threespine sticklebacks, *Gasterosteus aculeatus* (Pisces, Gasterosteidae). *Biological Journal of the Linnean Society* **31,** 347–382 (1987).

27. Bell, M. A. & Harris, E. I. Developmental osteology of the pelvic complex of *Gasterosteus aculeatus*. *Copeia* **1985,** 789 (1985).
28. Schmid, L. & Villagra, M. R. S. Potential genetic bases of morphological evolution in the triassic fish Saurichthys. *Journal of Experimental Zoology Part B: Molecular and Developmental Evolution* **314B,** 519–526 (2010).
29. Meredith, R. W., Gatesy, J., Murphy, W. J., Ryder, O. A. & Springer, M. S. Molecular decay of the tooth gene Enamelin (*ENAM*) mirrors the loss of enamel in the fossil record of placental mammals. *PLoS Genetics* **5,** e1000634 (2009).
30. Qu, Q., Haitina, T., Zhu, M. & Ahlberg, P. E. New genomic and fossil data illuminate the origin of enamel. *Nature* **526,** 108–111 (2015).
31. Zhu, M. *et al.* A Silurian placoderm with osteichthyan-like marginal jaw bones. *Nature* **502,** 188–193 (2013).
32. Zhu, M. Bone gain and loss: insights from genomes and fossils. *National Science Review* **1,** 490–497 (2014).
33. Hunt, G. Evolutionary divergence in directions of high phenotypic variance in the Ostracode genus *Poseidonamicus*. *Evolution* **61,** 1560–1576 (2007).
34. Thompson, J. R. *et al.* Reorganization of sea urchin gene regulatory networks at least 268 million years ago as revealed by oldest fossil cidaroid echinoid. *Scientific Reports* **5,** 15541 (2015).
35. Organ, C. L., Janes, D. E., Meade, A. & Pagel, M. Genotypic sex determination enabled adaptive radiations of extinct marine reptiles. *Nature* **461,** 389–392 (2009).
36. Organ, C. L., Shedlock, A. M., Meade, A., Pagel, M. & Edwards, S. V. Origin of avian genome size and structure in non-avian dinosaurs. *Nature* **446,** 180–184 (2007).
37. Organ, C. L. & Shedlock, A. M. Palaeogenomics of pterosaurs and the evolution of small genome size in flying vertebrates. *Biology Letters* **5,** 47–50 (2009).
38. Conte, G. L., Arnegard, M. E., Peichel, C. L. & Schluter, D. The probability of genetic parallelism and convergence in natural populations. *Proc. Biol. Sci.* **279,** 5039–5047 (2012).
39. Rennison, D. J., Stuart, Y. E., Bolnick, D. I. & Peichel, C. L. Ecological factors and morphological traits are associated with repeated genomic differentiation between lake and stream stickleback. *Philos. Trans. R. Soc. Lond., B, Biol. Sci.* **374,** 20180241 (2019).
40. Szeto, D. P. *et al.* Role of the Bicoid-related homeodomain factor Pitx1 in specifying hindlimb morphogenesis and pituitary development. *Genes Dev.* **13,** 484–494 (1999).
41. Thompson, A. C. *et al.* A novel enhancer near the Pitx1 gene influences development and evolution of pelvic appendages in vertebrates. *eLIFE* **7,** e38555 (2018).
42. Bell, M. A., Stewart, J. D. & Park, P. J. The world's oldest fossil Threespine Stickleback fish. *Copeia* **2009,** 256–265 (2009).
43. Rawlinson, S. E. & Bell, M. A. A stickleback fish (*Pungitius*) from the Neogene Sterling Formation, Kenai Peninsula, Alaska. *Journal of Paleontology* **56,** 583–588 (1982).
44. Rohlf, F. J. *tpsDIG ver. 2.10*. (2006).
45. Bowne, P. S. in *The Evolutionary Biology of the Threespine Stickleback* (eds. Bell, M. A. & Foster, S. A.) 28–60 (Oxford University Press, 1994).
46. R Core Team. R: A language and environment for statistical computing.
47. Lleonart, J., Salat, J. & Torres, G. J. Removing allometric effects of body size in Morphological analysis. *Journal of Theoretical Biology* **205,** 85–93 (2000).

48. Oke, K. B. *et al.* Does plasticity enhance or dampen phenotypic parallelism? A test with three lake-stream stickleback pairs. *J. Evol. Biol.* **29,** 126–143 (2016).
49. Stuart, Y. E. *et al.* Contrasting effects of environment and genetics generate a continuum of parallel evolution. *Nature Ecology and Evolution* **1,** 0158 (2017).
50. Hartigan, J. A. & Hartigan, P. M. The Dip Test of Unimodality. *The Annals of Statistics* **13,** 70–84 (1985).
51. Maechler, M. diptest: Hartigan's dip statistic for unimodality - corrected. version 0.75-7 (2016).

Acknowledgements: Cyprus Industrial Minerals, CR Minerals, and T. Sumner of World Minerals permitted us to collect fossils on their property. We thank the people acknowledged in [3] plus D. Arcieri, K. Brudvik, N.J. Buck, F. Castelli, R. Obeng, J. Qiao, C. Redman, and T.T. Zhang for field and laboratory assistance; F.J. Rohlf for statistical advice; M.D. Houseman for sharing his extensive knowledge of Truckee Formation paleoecology and taphonomy. C.W. Chan measured all pelvic vestiges for asymmetry analysis. We thank J. Jernvall, D.M. Kingsley, S. Chenoweth, R. Bonduriansky, Y.F. Chan, and S. Swank for helpful comments and discussion. This research was supported by NSF BSR-8111013, EAR-9870337, and DEB-0322818, Center for Field Research (Earthwatch), the National Geographic Society (2869-84), and NIH R01 GM124330-01 to M.A. Bell and by NSF DEB-1456462 and DEB-2003457 to Y.E. Stuart. This is contribution XXXX from Department of Biological Sciences, Rowan University.

Author contributions: MAB, MPT, and YES designed the research. MAB supervised sample and data collection. MPT collected the data. YES analyzed data, created figures, and wrote the paper. MAB, MPT, and YES jointly edited the paper.

Competing interests statement: We declare that none of the authors have competing financial or non-financial interests as defined by Nature Research.

Figures legends

Figure 1. The pelvic vestige was larger on the left side in significantly more than half of all specimens of *G. doryssus* with vestigial pelvic structures. The magnitude of asymmetry was greater when the left vestige was larger than when the right vestige was larger. Asymmetry of pelvic vestiges was calculated for 877 specimens from temporal sequence L. Each vertical bar shows asymmetry for one specimen. The vertical line represents zero asymmetry. Individuals to the left of the line have larger left vestiges.

Figure 2. Mean pelvic score declines through time in temporal sequence L after a delay. The last sample in which every fish had a pelvic score of 3.0 occurred 2,500 years after lineage II replaced lineage I (Table S2). Reduction accelerated about 3,750 years after replacement (Table S2). Means minus one standard error are shown.

Figure 3. Relative frequency distributions of pelvic score through time from temporal sequence K. Pelvic score is multimodal, suggesting Mendelian expression of *Pitx1* during pelvic reduction. Analysis and discussion in the main text describe the more finely resolved sampling of temporal sequence L. Samples from K are plotted here for ease of visualization. The patterns are qualitatively the same. k.T is the complete replacement of lineage I by lineage II. Time proceeds down the first column and then into the second column. Mean deposition time since the replacement event for each section is reported in years, as well as the mean pelvic score in the sample. Lines represent the proportion of specimens in each pelvic score category. Numbers above the lines are counts.

Figure 4. Reduction of size-adjusted pelvic girdle length (pgl) and pelvic spine length (psl) in temporal sequence K began immediately following replacement of lineage I by lineage II. In contrast, pelvic score, did not evolve substantially for another ~3750 years. Mean values are plotted with standard error bars. Arrows denote the first inferred breakpoint for each trait from piecewise regression. The slope of the first 'stick' is significantly negative for both pelvic girdle length and pelvic spine length. The slope of the first 'stick' for pelvic score is zero. Sample sizes are in Table S5.

Figure 5. Frequency distributions of (A) pelvic girdle length and (B) pelvic spine length for specimens with pelvic scores of 3.0 from temporal sequence K. Unimodality and normality suggest that multiple genes with additive effects underlie evolution in these traits. Time proceeds down. The oldest sample in this sequence is k.T, just after transition between lineage I and lineage II. Years since lineage II appeared are reported for each sample (Table S5), after which individuals with PS 3.0 become too rare to calculate a reasonable mean (Table S5). Black dots indicate sample means.

598 **Tables**

599 **Table 1.** Piecewise regression models using data from temporal sequence K confirm that
600 reduction of pelvic girdle length and pelvic spine length began in lineage II immediately after it
601 replaced lineage I. In contrast, pelvic score had a lag time of ~3750 years before reduction began.

(A) Pelvic girdle length	Coefficient estimate	Standard error	t-value	P
After-break intercept	11.68	0.118	98.60	< 0.001
Time†	-5.14×10^{-4}	3.3×10^{-5}	-15.7	0.001
Before-break*	1.09	0.163	6.7	0.006
Time x Before-break time‡	-1.34×10^{-3}	5.95×10^{-4}	-2.3	0.109

* Add to after-break intercept for before-break intercept: 11.68 + 1.09 = 12.77
† The slope of the after-break 'stick': -5.14×10^{-4}
‡ Add to after-break slope (†) for **before-break slope:** $-5.14 \times 10^{-4} + -1.34 \times 10^{-3} =$ **-1.86×10^{-3}**
Model significance: $F_{3,3} = 326.2$, P = 0.0003, Adj. $R^2 = 0.99$

(B) Pelvic spine length	Coefficient estimate	Standard error	t-value	P
After-break intercept	5.60	0.165	34.0	<0.001
Time†	-1.86×10^{-4}	4.54×10^{-5}	-4.1	0.026
Before break*	2.33	0.227	10.3	0.002
Time x Before-break time‡	-3.75×10^{-3}	8.27×10^{-4}	-4.5	0.020

* Add to global Intercept for before-break intercept: 5.60 + 2.33 = 7.93
† The slope of the after-break 'stick': -1.86×10^{-4}
‡ Add to after-break slope (†) for **before-break slope**: $-1.86 \times 10^{-4} + -3.75 \times 10^{-3} =$ **-3.94×10^{-3}**
Model significance: $F_{3,2} = 126.8$, P = 0.001, Adj. $R^2 = 0.98$

(C) Pelvic score	Coefficient estimate	Standard error	t-value	P
After-break intercept	4.67	0.00	8.8×10^{15}	< 0.001
Time†	-4.53×10^{-4}	0.00	-4.3×10^{14}	< 0.001
Before break *	-1.67	0.00	-3.1×10^{14}	<0.001
Time x Before-break time ‡	4.53×10^{-4}	0.00	3.9×10^{14}	<0.001

* Add to global Intercept for before-break intercept: 4.67 + -1.67 = 3.00
† The slope of the after-break 'stick': -4.53×10^{-4}
‡ Add to after-break slope (†) for **before-break slope**: $-4.53 \times 10^{-4} + 4.53 \times 10^{-4} =$ **0.00**
Model significance: $F_{3,3} = 13.53$, P < 0.001, Adj. $R^2 = 1$

602

603

Figures legends

Figure 1. The pelvic vestige was larger on the left side in significantly more than half of all specimens of *G. doryssus* with vestigial pelvic structures. The magnitude of asymmetry was greater when the left vestige was larger than when the right vestige was larger. Asymmetry of pelvic vestiges was calculated for 877 specimens from temporal sequence L. Each vertical bar shows asymmetry for one specimen. The vertical line represents zero asymmetry. Individuals to the left of the line have larger left vestiges.

Figure 2. Mean pelvic score declines through time in temporal sequence L after a delay. The last sample in which every fish had a pelvic score of 3.0 occurred 2,500 years after lineage II replaced lineage I (Table S2). Reduction accelerated about 3,750 years after replacement (Table S2). Means minus one standard error are shown.

Figure 3. Relative frequency distributions of pelvic score through time from temporal sequence K. Pelvic score is multimodal, suggesting Mendelian expression of *Pitx1* during pelvic reduction. Analysis and discussion in the main text describe the more finely resolved sampling of temporal sequence L. Samples from K are plotted here for ease of visualization. The patterns are qualitatively the same. k.T is the complete replacement of lineage I by lineage II. Time proceeds down the first column and then into the second column. Mean deposition time since the replacement event for each section is reported in years, as well as the mean pelvic score in the sample. Lines represent the proportion of specimens in each pelvic score category. Numbers above the lines are counts.

Figure 4. Reduction of size-adjusted pelvic girdle length (pgl) and pelvic spine length (psl) in temporal sequence K began immediately following replacement of lineage I by lineage II. In contrast, pelvic score, did not evolve substantially for another ~3750 years. Mean values are plotted with standard error bars. Arrows denote the first inferred breakpoint for each trait from piecewise regression. The slope of the first 'stick' is significantly negative for both pelvic girdle length and pelvic spine length. The slope of the first 'stick' for pelvic score is zero. Sample sizes are in Table S5.

Figure 5. Frequency distributions of (a) pelvic girdle length and (b) pelvic spine length for specimens with pelvic scores of 3.0 from temporal sequence K. Unimodality and normality suggest that multiple genes with additive effects underlie evolution in these traits. Time proceeds down. The oldest sample in this sequence is k.T, just after transition between lineage I and lineage II. Years since lineage II appeared are reported for each sample (Table S5), after which individuals with PS 3.0 become too rare to calculate a reasonable mean (Table S5). Black dots indicate sample means.