

## HUMAN GENETICS

# Adaptations to water stress and pastoralism in the Turkana of northwest Kenya

A. J. Lea<sup>1,2\*</sup>, I. V. Caldas<sup>3</sup>, K. M. Garske<sup>2,4,5</sup>, E. R. Gerlinger<sup>4,5</sup>, J. P. Arroyo<sup>6</sup>, J. Echwa<sup>2</sup>, M. Gurven<sup>7</sup>, C. Handley<sup>8</sup>, J. C. Kahumbu<sup>2,4,5</sup>, J. Kamau<sup>9,10</sup>, P. Kinyua<sup>2,11</sup>, F. Lotukoi<sup>2</sup>, A. Lopurdoi<sup>2</sup>, S. Lowasa<sup>2</sup>, S. N. Njeru<sup>11</sup>, R. Mallarino<sup>12</sup>, D. J. Martins<sup>2,13,14</sup>, P. W. Messer<sup>3</sup>, C. Miano<sup>2,5,11</sup>, B. Muhoya<sup>2,4,5</sup>, J. Peng<sup>4,5</sup>, T. Phung<sup>7,8</sup>, J. D. Rabinowitz<sup>5,15,16</sup>, A. Roichman<sup>5,15</sup>, R. Siford<sup>8</sup>, A. C. Stone<sup>8,17,18</sup>, A. M. Taravella Oill<sup>17,18</sup>, S. Mathew<sup>8,17</sup>, M. A. Wilson<sup>18,19</sup>, J. F. Ayroles<sup>2,21\*</sup>

The Turkana pastoralists of Kenya inhabit arid, water-limited environments and rely largely on livestock for subsistence. Working with Turkana communities, we sequenced 367 whole genomes and identified eight regions with evidence for recent positive selection. One of these regions includes a putative regulatory element for *STC1*—a kidney-expressed gene involved in metabolism and the response to dehydration. We show that *STC1* is induced by antidiuretic hormone in human cells, is associated with urea levels in the Turkana themselves, and is under strong and recent selection in this population as well as a second East African population, the Daasanach. This work highlights how integrating anthropological and genomic approaches can lead to a new understanding of human physiology with biomedical relevance.

Humans inhabit an astounding variety of challenging habitats—from high-altitude plateaus to arctic tundras to scorching deserts. In many cases, natural selection has played a key role in sustaining human populations in extreme environments (1, 2). For example, the hypoxic conditions experienced by Tibetans living at high altitudes have selected for changes in red blood cell production regulated by *EPAS1* (3), whereas the seafood-centered diets of Greenlandic Inuit have selected for changes in fatty acid metabolism regulated by the *FADS* gene cluster (4). Such examples highlight that, in addition to helping us understand our evolutionary history, studies of natural selection can uncover new genotype-phenotype links of biomedical importance. However, despite great interest in uncovering the genetic basis of adaptation, few studies have robustly linked ecological selection pressures, genetic variation, and human phenotypes (2).

To do so, we worked with the Turkana people of northwest Kenya, who live in environments that present several dietary and climatic challenges (Fig. 1, A and B). The Turkana are an Eastern Nilotic group, and like other members of their lineage, they originated in the Nile Valley region and began nomadic pastoralist practices ~5000 to ~8000 years ago (5, 6). Oral histories suggest that the Turkana migrated from the Karamoja plateau highlands of Uganda to the more arid rift valley area of northwest Kenya

200 to 250 years ago. They managed the drier environment by further intensifying their reliance on livestock compared with other linguistically similar groups in the Karamojong cluster (7–9).

Although present-day Turkana pastoralists consume agricultural products obtained via trade and small-scale markets (e.g., maize, wheat flour, legumes, and sugar), the Turkana diet is overall rich in protein: previous studies have estimated that in certain areas and seasons, 70 to 80% of the diet is animal-derived (62% from milk and 8 to 18% from blood, marrow, and red meat) (10). Our dietary interviews with Turkana pastoralists revealed that 74% consumed blood several times per week, and 96.8% of individuals ate red meat at least once per week ( $n = 346$ ; Fig. 1C and table S1).

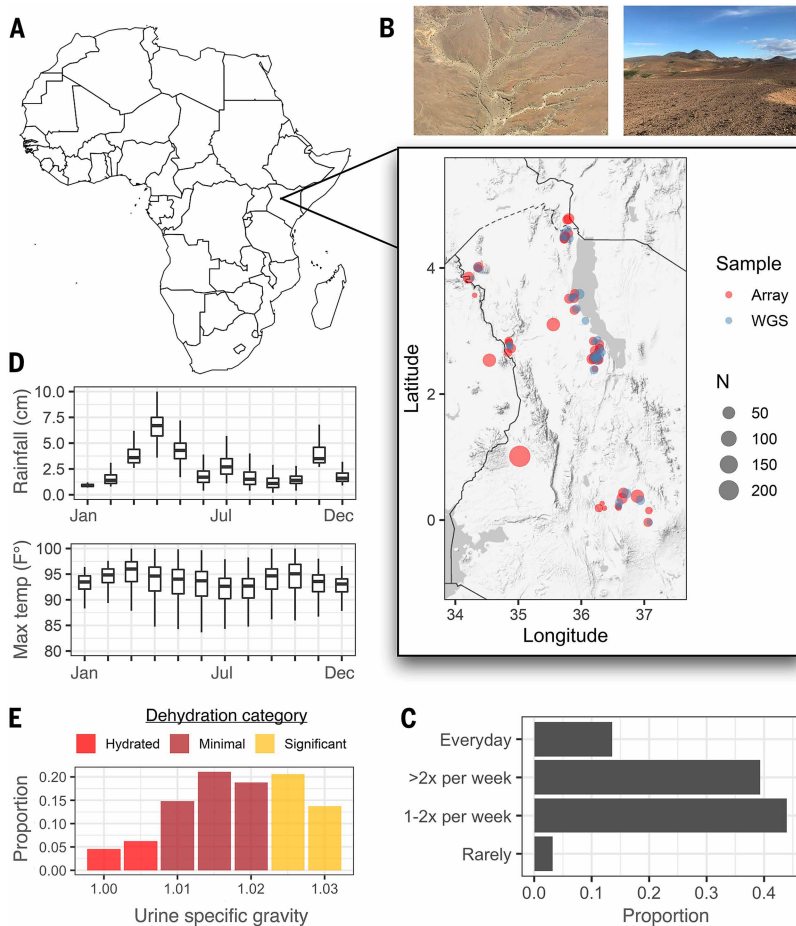
The Turkana practice nomadic pastoralism because they inhabit extremely arid, water-limited (Fig. 1D) landscapes that cannot easily support agriculture, hunting and gathering, or other common subsistence strategies; instead, they effectively access transient rainfall and vegetation used by their livestock. To quantify the impact of water limitation, we interviewed Turkana pastoralists and found that water stress was a daily issue: 76% of Turkana spend more than a few hours a day collecting water, and 99% perceive this water to be insufficient in amount ( $n = 311$ ; table S1). Using measures of urine-specific gravity, we also found that 89% of people meet the physiological criteria (11) for minimal (55%) or significant (34%) dehydration ( $n = 175$ ; Fig. 1E and table S1), emphasizing the challenge of maintaining fluid balance in this environment.

## Evidence of selection in Turkana genomes

We hypothesized that exposure to an animal product-rich diet and an arid ecology would select for genetic variants regulating metabolism and dehydration stress in the Turkana. To test this, we scanned for signatures of selection in 308 Turkana genomes sequenced at high (>20 $\times$ ,  $n = 106$ ) and medium (~6 $\times$ ,  $n = 202$ ) coverage. To understand population genetic parameters that could affect our analyses and interpretation, we also sequenced 59 genomes from nearby Kenyan and Ugandan groups, namely the El Molo, Ik, Karamojong, Masaai, Ngitepes (also known as Tepeth), Pokot, Rendille, and Samburu (fig. S1 and tables S2 and S3).

We genotyped 7,767,165 variants with a minor allele frequency >1% in our high-coverage Turkana dataset and then imputed missing data in the medium-coverage samples (Fig. 2A and fig. S2). We performed extensive quality control on our combined, imputed whole-genome sequencing (WGS) dataset (figs. S3 to S6), including corroborating our WGS-derived genotype calls using independent calls from the Infinium Global Screening Array ( $R^2$  between WGS- and array-derived genotypes for 108 paired samples =  $0.96 \pm 0.03$ ; table S4). Overall, population genetic analyses of the WGS dataset highlighted two key takeaways that were confirmed with the array dataset ( $n = 783$  individuals total) and are consistent with the literature (6, 12). First, there has been little European admixture in the East African groups we worked with (mean European ancestry proportion estimated by RFMix =  $6.6 \pm 4.4\%$ ; figs. S7 to S9 and table S5). Second, although they are culturally and linguistically distinct, the Turkana are similar at the genomic level to the Ik, Karamojong, Masaai, Ngitepes, and Pokot when examining common summary statistics (e.g., genotype PC loadings or  $F_{ST}$ ; Fig. 2A; figs. S5, S8, and S10; and table S6).

<sup>1</sup>Department of Biological Sciences, Vanderbilt University, Nashville, TN, USA. <sup>2</sup>Turkana Health and Genomics Project, Turkana Basin Institute, Nairobi, Kenya. <sup>3</sup>Department of Computational Biology, Cornell University, Ithaca, NY, USA. <sup>4</sup>Department of Ecology and Evolutionary Biology, Princeton University, Princeton, NJ, USA. <sup>5</sup>Lewis Sigler Institute for Integrative Genomics, Princeton University, Princeton, NJ, USA. <sup>6</sup>Division of Nephrology and Hypertension, Department of Medicine, Vanderbilt University Medical Center, Nashville, TN, USA. <sup>7</sup>Department of Anthropology, University of California Santa Barbara, Santa Barbara, CA, USA. <sup>8</sup>School of Human Evolution and Social Change, Arizona State University, Tempe, AZ, USA. <sup>9</sup>Institute of Primate Research, National Museums of Kenya, Nairobi, Kenya. <sup>10</sup>Department of Biochemistry, School of Medicine, University of Nairobi, Nairobi, Kenya. <sup>11</sup>Center for Community Driven Research, Kenya Medical Research Institute, Kirinyaga, Kenya. <sup>12</sup>Department of Molecular Biology, Princeton University, Princeton, NJ, USA. <sup>13</sup>Turkana Basin Institute, Turkana, Kenya. <sup>14</sup>Department of Anthropology, Stony Brook University, Stony Brook, NY, USA. <sup>15</sup>Department of Chemistry, Princeton University, Princeton, NJ, USA. <sup>16</sup>Ludwig Institute for Cancer Research, Princeton University, Princeton, NJ, USA. <sup>17</sup>Institute of Human Origins, Arizona State University, Tempe, AZ, USA. <sup>18</sup>Center for Evolution and Medicine, Arizona State University, Tempe, AZ, USA. <sup>19</sup>School of Life Sciences, Arizona State University, Tempe, AZ, USA. <sup>20</sup>Comparative Genomics and Reproductive Health Section, Center for Genomics and Data Science Research, National Human Genome Research Institute, National Institutes of Health, Bethesda, MD, USA. <sup>21</sup>Department of Integrative Biology, University of California, Berkeley, Berkeley, CA, USA. \*Corresponding author. Email: amanda.j.lea@vanderbilt.edu (A.J.L.); ayroles@berkeley.edu (J.F.A.)



**Fig. 1. Ecology and lifestyle of the Turkana people.** (A) Map of Africa with an inset showing northwest Kenya (the present-day homelands of the Turkana people). Dots indicate where samples were collected for this study, for both WGS and array genotyping. (B) Representative photographs of the arid ecology of the Turkana region (photos taken by the authors). (C) Proportion of Turkana pastoralists in this study who consume meat at different self-reported frequencies ( $N = 346$ ). (D) Average rainfall and maximum temperature for the Turkana country region by month. Data were sourced from WorldClim (56). (E) Proportion of Turkana pastoralists with different urine specific gravity values ( $N = 175$ ), colored by whether each value meets the criteria for dehydration provided in (11).

To identify potential selective sweeps in the Turkana, we computed three statistics that rely on different assumptions and subsets of the data and that detect selection on the order of thousands to tens of thousands of years ago (2): (i) the integrated haplotype score (iHS) (13), computed on high-coverage Turkana genomes only; (ii) the population branch statistic (PBS) (3), computed on high- and medium-coverage Turkana genomes; and (iii) the XiX statistic (14), computed on all genomes. Additionally, our PBS and XiX analyses included the Luhya (East Africa) and Yoruba (West Africa) populations from the 1000 Genomes Project as outgroups ( $n = 207$ ) (15). After computing all three statistics, we used a sliding-window approach to intersect the results and identified 13 50-kb outlier windows near eight genes (Fig. 2B and table S7). Two of these eight genes (*CCDC102B* and *SEMA6A*) were previously found to be under selection in the Maasai, a closely related Nilotic pastoralist group (16). To our knowledge, the rest have not been previously identified as targets of selection in humans.

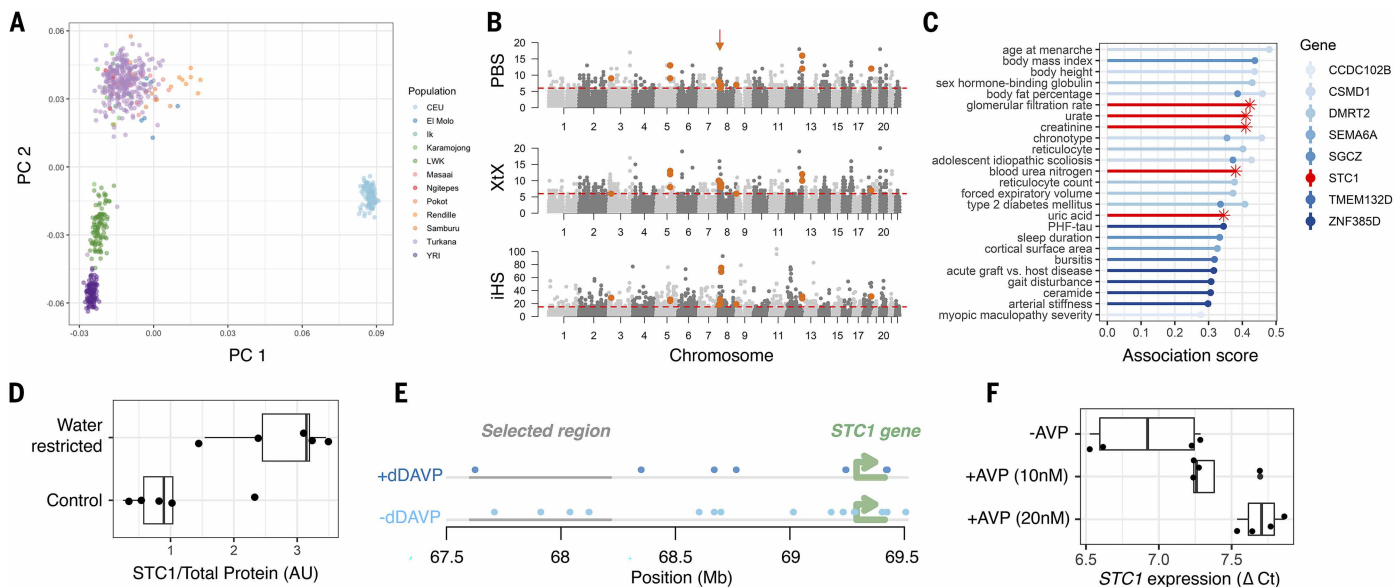
When we investigated what phenotypes our candidate genes have been previously linked to by genome-wide association studies (GWASs), which have primarily focused on European ancestry cohorts, we found that they were enriched for a major predictor of cardiovascular disease—arterial stiffness [Fisher's exact test, odds

ratio = 16.70, false discovery rate (FDR)-adjusted  $P$  value = 0.047]. Unexpectedly, they were also enriched for neurological biomarkers of Alzheimer's disease—namely, neurofibrillary tangles (odds ratio = 19.76, FDR-adjusted  $P$  value =  $3.24 \times 10^{-4}$ ), PHF-tau (odds ratio = 5.68, FDR-adjusted  $P$  value = 0.079), and cortical surface area (odds ratio = 6.52, FDR-adjusted  $P$  value = 0.079; Fig. 2C and tables S8 and S9). In line with this observation, several of our candidate genes are most highly expressed in neurological cell types, such as inhibitory neurons, oligodendrocytes, and oligodendrocyte precursors (table S10).

### Selection on *STC1*, a regulator of metabolic and renal system functions

Some of the strongest biological evidence from previous studies pointed to *STC1*, a gene that encodes a glycoprotein with autocrine and paracrine functions. Specifically, previous GWASs have linked *STC1* to (i) serum levels of urate, a waste product produced when the body breaks down purine-rich foods such as red meat; (ii) serum levels of urea and creatinine, two common biomarkers of kidney function; and (iii) estimated glomerular filtration rate, a measure of kidney health (17) (Fig. 2C and table S8). Beyond GWASs, *STC1* has been implicated in adaptation to challenging environments in multiple natural animal populations (18–20), and laboratory-based model organism studies have implicated *STC1* in the ability to suppress reactive oxygen species and reduce acute kidney injury (21, 22), glucose homeostasis (23), and the response to dehydration. In particular, *STC1* transcription is induced up to eightfold in rodent kidneys after water deprivation (24), a response that is coordinated by antidiuretic hormone (ADH) (25) and involves *STC1*-based regulation of both rising hypertonicity and progressive hypovolemia (26). We replicated these results at the protein level, showing that *STC1* levels are higher in water-restricted versus control mouse kidneys ( $t$  test,  $P$  value =  $6.74 \times 10^{-3}$ ; Fig. 2D and table S12). Given the clear involvement of *STC1* in metabolic and renal system traits of ecological relevance to the Turkana people, we prioritized this gene for follow-up analyses.

We identified two overlapping 50-kb candidate regions near *STC1*, with the collapsed 75-kb region located ~150 kb upstream of the transcription start site. This region was ranked first out of all tested regions by the iHS statistic and third by the PBS and XiX statistics (fig. S11 and table S7). Previously published high-throughput chromosome conformation capture (Hi-C) data from five tissues that express *STC1* (27) show that this regulatory region and the *STC1* gene body fall within the same topological domain (28) and are in consistent contact (29) (table S11, fig. S12). In our main tissue of interest, the kidney, *STC1* is highly specific to cell type and is expressed almost exclusively in the collecting duct (30, 31). This structure is the final segment of the kidney to control fluid balance, accounting for ~5% of water reabsorption at baseline and up to ~25% during ADH surges induced by dehydration. When we analyzed ATAC-seq data previously generated from a mouse collecting duct cell line exposed versus unexposed to ADH (32), we found that our candidate region contained several differentially accessible, ADH-responsive regulatory elements (Fig. 2E). Although there is no human collecting duct cell line, we performed new experiments to show that *STC1* is induced by ADH in a human kidney (epithelial) cell line (linear model,  $P$  value for 10 and 25 nM = 0.033 and 0.0042, respectively; Fig. 2F and table S12).



**Fig. 2. Scans for selection and evidence for the *STC1* gene.** (A) Principal components (PC) analysis comparing WGS data from the Turkana as well as other study communities to African populations included in the 1000 Genomes Study (15). CEU, Northern Europeans from Utah; LWK, Luhya in Kenya; YRI, Yoruba in Nigeria. (B) Number of PBS, XtX, and iHS outliers per 50-kb window. Red dotted lines represent the cutoff for the 99th percentile of each empirical distribution for each selection statistic. Windows that exceeded the 99th percentile for all three statistics are highlighted in red. The candidate region that falls near the *STC1* gene is highlighted with a red arrow. (C) Phenotypes associated with genes that fall in or near candidate regions. Phenotypic associations were sourced from the Open Targets Platform (57), and the association score (*y* axis) represents the aggregate evidence across all published studies for a gene-phenotype link. Only association scores >0.25 are plotted, and association scores involving the *STC1* gene are highlighted with an asterisk. (D) Normalized *STC1* protein concentrations in mouse kidneys after 24 hours of water restriction. (E) Open chromatin regions in mouse collecting duct cells exposed versus unexposed to desmopressin (dDAVP) (synthetic analog of vasopressin) (32). Coordinates are in mm10; open chromatin regions are highlighted in blue, and the *STC1* candidate region is highlighted in gray. (F) Quantitative polymerase chain reaction of *STC1* expression levels in human embryonic kidney 293 cells exposed versus unexposed to ADH (equivalent to AVP).

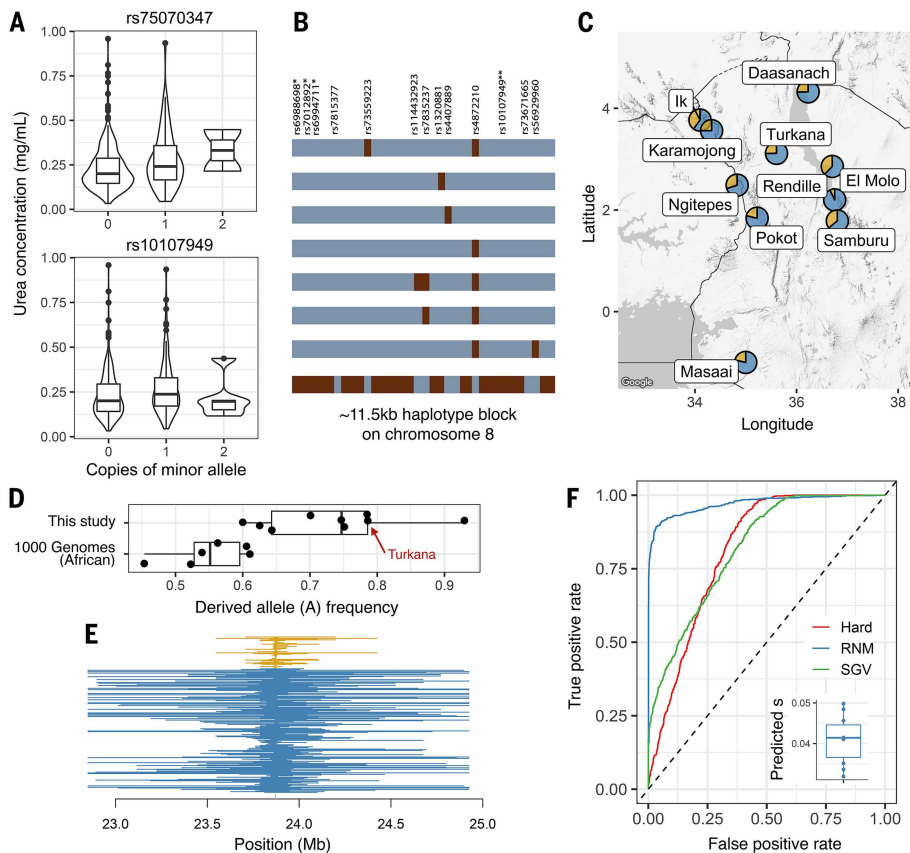
To further link our candidate region to kidney function, we measured serum creatinine and urea levels for 447 Turkana included in our genotyping array dataset. After subsetting to single-nucleotide polymorphisms (SNPs) within the *STC1* candidate region that passed our filters, we were able to test six SNPs typed on the array for associations with these outcomes; SNP density in the array dataset is sparse, and these six SNPs are thus expected to tag general haplotype structure (fig. S13). Using this approach, we found two SNPs that were significantly correlated with urea levels [linear mixed effects model, rs10107949 beta = 0.030 (FDR-adjusted *P* value = 0.059), rs75070347 beta = 0.040 (FDR-adjusted *P* value = 0.055); Fig. 3A and tables S13 and S14]. One of these SNPs falls on the same ~11.5-kb haplotype block as the top three derived SNPs with the strongest evidence for selection in the WGS dataset (rs6988698, rs7012892, and rs6994711; fig. S14). This haplotype block also contains several candidate regulatory elements in kidney [using ENCODE data and annotations (33); fig. S15]. Although further work is needed to identify the causal variant in the *STC1* region and the precise mechanism of action, we propose that these data, in aggregate, point toward selection on *STC1* regulation in the context of ADH induction, dehydration stress, and kidney function.

### Evolutionary history of *STC1*

We next turned our attention toward understanding the evolutionary history of the *STC1* regulatory region, including the nature, strength, and timing of selection. We started by checking the worldwide allele frequencies of the top three derived SNPs with the strongest evidence for selection in the WGS dataset (derived allele frequency in Turkana for rs6988698 is 75%; rs7012892, 84%; and rs6994711, 84%). We found that these same SNPs are near invariant outside of Africa, with average

minor allele frequencies of 2.66% (range = 0 to 8.65%; figs. S16 and S17 and table S15); for these three SNPs, the derived alleles are at high frequency outside of Africa, potentially resulting from genetic drift after the out-of-Africa bottleneck and highlighting the difficulty of interpreting selection signals in African populations with complex demographic histories. Within Africa, these three example variants have intermediate derived allele frequencies in non-Nilotic East African groups and at Turkana-like frequencies in the other East African groups we sampled (Fig. 3, C to E). The East African groups are almost all part of the same Nilotic lineage the Turkana belong to, which began pastoralist practices ~5000 to ~8000 years ago (5, 6). Thus, we hypothesize that there was selection on standing variation within the Nilotic cluster starting sometime after pastoralist practices emerged.

To test this scenario, we estimated the site frequency spectrum for the Turkana and inferred their demographic history (figs. S18 and S19 and table S16). We then simulated genomic datasets under different evolutionary scenarios—varying the nature, strength, and timing of selection in the *STC1* region. We trained two separate convolutional neural networks (CNNs) with the same data: a regression model to infer sweep strength and a classification model to classify between three potential modes of selection (hard sweep, selection on recurrent new mutations, and selection on standing genetic variation) (34) (fig. S20). When applied to our real data, the CNNs inferred that selection on standing variation began ~348 generations ago (range, 187 to 571), with a selection coefficient of ~0.041 (range, 0.033 to 0.05; Fig. 3F and fig. S21). These estimates should be interpreted with caution, as they depend on assumptions about local demography and the unknown initial frequency of the selected alleles—parameters that remain poorly resolved. However, even under wide uncertainty, our



**Fig. 3. Selection on genetic variation near the *STCI* gene.** (A) Correlations between Turkana genotypes within the *STCI* region and serum urea levels (rs10107949:  $R = 0.116$ , FDR-adjusted  $P$  value = 0.059; rs75070347:  $R = 0.130$ , FDR-adjusted  $P$  value = 0.055). (B) Haplotype block 21, which is 11.5 kb (spanning chr8:23872028–23883577) and contains the three SNPs with the strongest evidence for selection (rs6988698, rs7012892, and rs6994711), as well as one SNP, rs10107949, which was found to be associated with urea levels in the Turkana. Dark colors indicate minor alleles, and light colors indicate major alleles. (C) Allele frequency for one of the strongest evidence SNPs (rs6988698) within data generated as part of this study. The map shows the frequency of the major or derived (blue) versus minor or ancestral (yellow) alleles. (D) Boxplot summarizing the frequency of rs6988698 in this study as well as African populations included in the 1000 Genomes Project (15). (E) Haplotype length at the rs6988698 SNP for individuals carrying the major or derived (blue) versus minor or ancestral (yellow) allele. Coordinates are in hg19. (F) Validation of the CNN used to infer sweep mode. Receiver operating characteristic (ROC) curves show the results of applying one CNN replicate on the simulated validation dataset. Each curve represents a one-versus-all comparison between the reference mode and all others combined; the area under each curve is given in parentheses. Hard, hard sweep [area under the ROC curve (AUC), 0.813; RNM, selection on recurrent new mutations (AUC, 0.971); SGV, selection on standing genetic variation (AUC, 0.819)]. Inset shows the results of applying 10 replicate CNNs to real data from the *STCI* locus, under the assumption that the selective sweep is codominant (see fig. S20 for results when the sweep is assumed to be dominant).

analyses suggest that selection was probably strong; for example, selection coefficient estimates for the malaria-protective Duffy (35) and lactase persistence (36, 37) alleles range from 0.04 to 0.08 and 0.04 to 0.06, respectively. Our analyses also point toward an onset of selection sometime within the past ~7000 years. This timing overlaps with increasing aridification across East Africa (38) and the emergence of pastoralism in the region (39, 40).

### Selection on *STCI* in a Cushitic population, the Daasanach

Our CNN analyses indicated that the *STCI* haplotype of interest was likely segregating in East Africa before the Nilotic lineage split with the Eastern Cushitic lineage (speaking Afro-Asiatic rather than Nilo-Saharan languages). Both lineages are thought to have originated in the Nile Valley region but diverged during their distinct migrations to present-day Kenya and Ethiopia (41–44). Thus, we wondered whether

there was also evidence for selection on *STCI* within this separate lineage, which has similarly inhabited arid regions around the horn of Africa and practiced a pastoralist lifestyle (45, 46). To do so, we worked with Daasanach communities on the east side of Lake Turkana, where ongoing work suggests a complex, and potentially compensatory, relationship between water stress and hydration status (47–49). We generated WGS data from 95 Daasanach participants and tested for signatures of selection on chromosome 8 (which contains *STCI*). We found that the same region upstream of *STCI* identified in the Turkana is also an iHS outlier in the Daasanach (fig. S22 and table S17). This finding suggests that the regulatory region upstream of *STCI* has been under selection in both lineages.

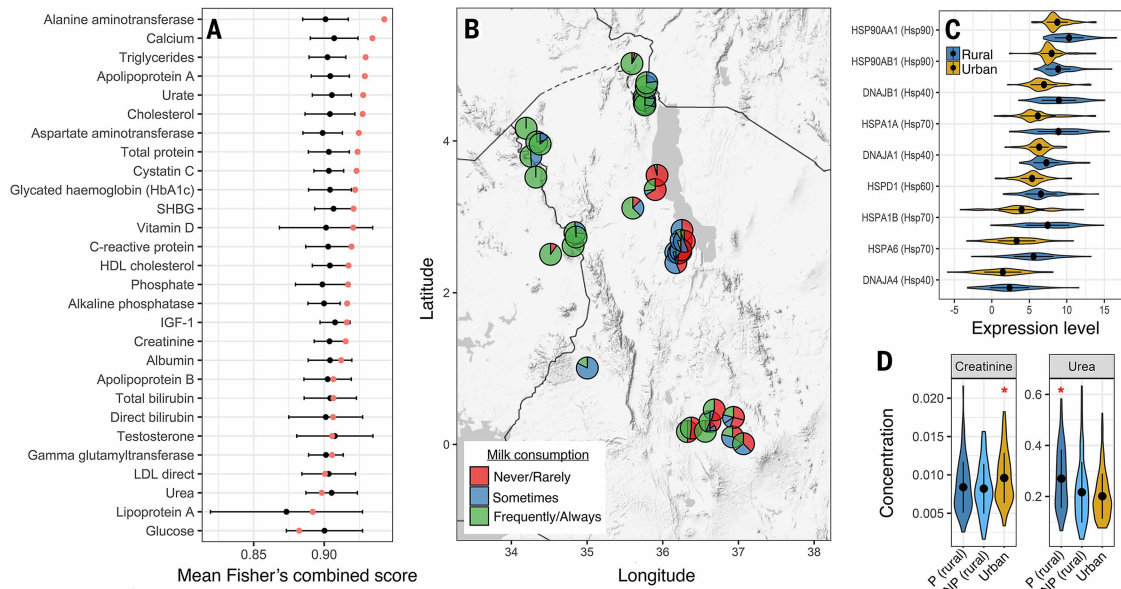
### Polygenic selection on metabolic and renal system biomarkers

Although selective sweeps—such as the one we have detected near *STCI*—are undoubtedly important in human evolution (2), complex traits are more commonly shaped by weak but simultaneous selection on numerous genetic variants (50). We were therefore motivated to also explore the contribution of polygenic adaptation to metabolic and renal system traits in the Turkana. To do so, we collapsed our three selection statistics into a Fisher's combined score (FCS) and assessed whether regions previously associated with 29 biomarkers (via multi-ancestry GWASs in the UK Biobank) exhibited higher mean FCS scores in the Turkana than expected by chance (51). These analyses revealed evidence of polygenic selection on metabolic biomarkers, such as triglycerides, cholesterol, and HbA1c, as well as renal system biomarkers, such as uric acid and cystatin C (all FDR < 5%; Fig. 4A and table S18). These results suggest that there has been diffuse, genome-wide selection on metabolism and kidney function in the Turkana, in addition to the localized signal we uncovered at *STCI*.

### Lifestyle change and the fate of selected alleles

Although many Turkana still practice nomadic pastoralism in northwest Kenya and thus experience the ecological selection pressures we have highlighted here, subsets of the population are undergoing different types and degrees of market integration, urbanization, and acculturation. In recent decades, some Turkana have stopped practicing pastoralism but continue to live in rural environments in northwest Kenya, whereas others have migrated to urban areas in favor of wage labor jobs. Not surprisingly, moving to a city represents a major environmental shift: for example, whereas >90% of pastoralists regularly consume blood, milk, and red meat, these numbers drop to 0, 47, and 31%, in urban areas where large portions of the diet are instead market-derived (52, 53) (Fig. 4B).

This ongoing lifestyle shift is relevant to our evolutionary genomic analyses because it sets the Turkana up for “evolutionary mismatch” (54). This occurs when previously advantageous variants, selected for in past ecologies, are placed in novel environments where they instead



**Fig. 4. Polygenic selection and lifestyle change.** (A) Red dots represent the mean FCS (representing a composite of all three selection statistics) for each 50-kb window that contained a SNP associated with the focal trait (associations were sourced from multi-ancestry GWASs in the UK Biobank). Boxplots show the FCS distribution for 10,000 randomly sampled windows matched for SNP density, recombination rate, and background selection. HDL, high-density lipoprotein; LDL, low-density lipoprotein; SHBG, sex hormone-binding globulin. (B) Self-reported milk consumption summarized by location (for Turkana individuals only). (C) Normalized gene expression levels for samples collected from Turkana individuals in rural versus urban locations. Because of small sample sizes, the rural category includes pastoralist and nonpastoralist individuals (52); results are qualitatively similar when they are split (fig. S24). Plot includes all heat shock proteins in the “response to heat” gene ontology category that were significantly differentially expressed at a 5% FDR. (D) Serum concentrations of urea (mg/mL) and creatinine (mg/mL) for rural pastoralists (P), rural nonpastoralists (NP), and urban individuals. Red asterisks indicate groups that were significantly different from the other two (FDR < 5%).

have detrimental effects. In line with this idea, we have previously shown that transitions from pastoralist to urban lifestyles are associated with increases in biomarkers of cardiovascular disease risk (52). We generated new data on (i) serum urea and creatinine levels and found that both kidney function biomarkers also differ between urban and rural-pastoralist settings ( $n = 447$ , all FDR < 5%; Fig. 4D and table S14), as well as (ii) blood gene expression levels and found that genes involved in biological processes such as “response to heat,” “protein folding,” and “inflammatory response” are differentially regulated between urban and rural-pastoralist individuals ( $n = 230$ , all FDR < 5%; Fig. 4C, fig. S23, and tables S19 and S20). Furthermore, we found that SNPs that fall near genes differentially expressed by lifestyle exhibit more evidence for selection than SNPs near genes unaffected by lifestyle change (linear model  $P$  values:  $iHS < 2 \times 10^{-16}$ ,  $PBS = 0.007$ , and  $XtX = 0.0211$ ; tables S21 and S22). This result suggests that past adaptations are poised to interact with the environmental and physiological shifts the Turkana are currently experiencing.

## Conclusions

We integrated anthropological, biological, and genomic datasets to show that an arid ecology combined with pastoralist practices has likely led to selection on *STC1* in the Turkana; this signal appears to be shared across other groups in the region, as emphasized by our work with the Daasanach. Although we do not know the causal, selected allele at this time or the specifics of when and how the allele underwent selection, we show that the *STC1* gene is induced by ADH in human cells, that *STC1* variants are linked to urea levels in the Turkana themselves, and that selection likely occurred on standing variation near *STC1* on an ecologically plausible time scale. We also shed rare empirical light on the popular evolutionary mismatch hypothesis, which is commonly invoked to explain the high rates of non-communicable, “lifestyle” diseases observed in Western countries (55). Our work highlights that partnerships with transitioning populations

can lead to new models for understanding how present-day environments interact with past adaptations to influence disease risk.

## REFERENCES AND NOTES

1. M. L. Benton *et al.*, *Nat. Rev. Genet.* **22**, 269–283 (2021).
2. S. Fan, M. E. B. Hansen, Y. Lo, S. A. Tishkoff, *Science* **354**, 54–59 (2016).
3. X. Yi *et al.*, *Science* **329**, 75–78 (2010).
4. M. Fumagalli *et al.*, *Science* **349**, 1343–1347 (2015).
5. O. Hanotte *et al.*, *Science* **296**, 336–339 (2002).
6. P. Triska *et al.*, *Genome Biol. Evol.* **7**, 3484–3495 (2015).
7. B. L. Korobe, *J. Afr. Interdiscip. Stud.* **7**, 29–67 (2023).
8. P. H. Gulliver, *The Family Herds: A Study of Two Pastoral Tribes in East Africa, The Jie and Turkana* (Routledge, 2013).
9. J. Lamphear, *J. Afr. Hist.* **29**, 27–39 (1988).
10. K. A. Galvin, *Am. J. Hum. Biol.* **4**, 209–221 (1992).
11. D. J. Casa *et al.*, *J. Athl. Train.* **35**, 212–224 (2000).
12. S. A. Tishkoff *et al.*, *Science* **324**, 1035–1044 (2009).
13. B. F. Voight, S. Kudravalli, X. Wen, J. K. Pritchard, *PLoS Biol.* **4**, e72 (2006).
14. T. Günther, G. Coop, *Genetics* **195**, 205–220 (2013).
15. The 1000 Genomes Project Consortium, *Nature* **491**, 56–65 (2012).
16. K. Wagh *et al.*, *PLoS ONE* **7**, e44751 (2012).
17. G. B. Loeb *et al.*, *Nat. Genet.* **56**, 2078–2092 (2024).
18. C.-L. Zhang *et al.*, *Sci. Rep.* **12**, 12264 (2022).
19. S. A. Rahmatalla *et al.*, *BMC Genet.* **18**, 92 (2017).
20. L. Kordonowy, M. MacManes, *BMC Genomics* **18**, 473 (2017).
21. L. Huang *et al.*, *Kidney Int.* **82**, 867–877 (2012).
22. J. S.-C. Pan *et al.*, *J. Am. Soc. Nephrol.* **26**, 364–378 (2015).
23. V. Schein *et al.*, *Mol. Cell. Endocrinol.* **414**, 1–8 (2015).
24. J. Turner, O. Sazonova, H. Wang, A. Pozzi, G. F. Wagner, *Mol. Cell. Endocrinol.* **328**, 8–15 (2010).
25. A. Y. Law *et al.*, *Mol. Cell. Endocrinol.* **348**, 183–188 (2012).
26. J. Turner, F.-L. Xiang, Q. Feng, G. F. Wagner, *Mol. Cell. Endocrinol.* **331**, 150–157 (2011).
27. A. Battle *et al.*, *Nature* **550**, 204–213 (2017).
28. J. R. Dixon *et al.*, *Nature* **485**, 376–380 (2012).
29. Y. Wang *et al.*, *Genome Biol.* **19**, 151 (2018).
30. J. Park *et al.*, *Science* **360**, 758–763 (2018).
31. Y. Muto *et al.*, *Nat. Commun.* **12**, 2190 (2021).

32. H. J. Jung, V. Raghuram, J. W. Lee, M. A. Knepper, *J. Am. Soc. Nephrol.* **29**, 1490–1500 (2018).
33. J. E. Moore *et al.*, *Nature* **583**, 699–710 (2020).
34. I. V. Caldas, A. G. Clark, P. W. Messer, bioRxiv 2022.07.19.500702 [Preprint] (2022). <https://doi.org/10.1101/2022.07.19.500702>
35. I. Hamid, K. L. Korunes, S. Bezeza, A. Goldberg, *eLife* **10**, e63177 (2021).
36. J. Burger *et al.*, *Curr. Biol.* **30**, 4307–4315.e13 (2020).
37. L. Ségurel, C. Bon, *Annu. Rev. Genomics Hum. Genet.* **18**, 297–319 (2017).
38. J. E. Tierney, P. B. deMenocal, *Science* **342**, 843–846 (2013).
39. E. A. Hildebrand *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **115**, 8942–8947 (2018).
40. F. Marshall, E. Hildebrand, *J. World Prehist.* **16**, 99–143 (2002).
41. A. J. Ahmed, Ed., *The Invention of Somalia* (Red Sea Press, 1995).
42. C. Ehret, *Southern Nilotic History* (Northwestern Univ. Press, 2019).
43. B. M. Lynch, L. H. Robbins, *J. Afr. Hist.* **20**, 319–328 (1979).
44. W. S. Nasong'o, M. N. Amutabi, T. Falola, Eds., *The Palgrave Handbook of Kenyan History* (Springer International Publishing, ed. 1, 2024).
45. E. S. Poloni *et al.*, *Ann. Hum. Genet.* **73**, 582–600 (2009).
46. R. Blench, in *L'Homme et l'animal. Dans le bassin du lac Tchad*, C. Baroin, J. Boutrais, Eds. (Institut de recherche pour le développement, 1999), pp. 39–80.
47. L. Ford *et al.*, *Water Int.* **48**, 63–86 (2023).
48. A. Y. Rosinger *et al.*, *Am. J. Hum. Biol.* **34**, e23715 (2022).
49. H. J. Bethancourt *et al.*, *Am. J. Hum. Biol.* **33**, e23447 (2021).
50. E. A. Boyle, Y. I. Li, J. K. Pritchard, *Cell* **169**, 1177–1186 (2017).
51. M. Lopez *et al.*, *Curr. Biol.* **29**, 2926–2935.e4 (2019).
52. A. J. Lea, D. Martins, J. Kamau, M. Gurven, J. F. Ayroles, *Sci. Adv.* **6**, eabb1430 (2020).
53. A. J. Lea *et al.*, *Evol. Med. Public Health* **9**, 406–419 (2021).
54. D. E. Lieberman, *The Story of the Human Body: Evolution, Health, and Disease* (2013).
55. A. J. Lea *et al.*, *PLoS Biol.* **21**, e3002311 (2023).
56. S. E. Fick, R. J. Hijmans, *Int. J. Climatol.* **37**, 4302–4315 (2017).
57. D. Ochoa *et al.*, *Nucleic Acids Res.* **49**, D1302–D1310 (2021).
58. A. Lea, election\_19May25\_v2, Version v1.0.1, Zenodo (2025); <https://doi.org/10.5281/zenodo.15468316>.
59. A. Lea, Genotype calls from whole genome sequencing: Turkana high coverage samples, Zenodo (2025); <https://doi.org/10.5281/zenodo.16588211>.
60. A. Lea, Genotype calls from whole genome sequencing: Turkana high coverage samples, Zenodo (2025); <https://doi.org/10.5281/zenodo.16556999>.

## ACKNOWLEDGMENTS

We thank previous members of the Turkana Health and Genomics Project for their contributions, especially S. Lowasa, D. Mukhongo, S. Ngatia, E. Loowoth, and M. Ndegwa. We are also grateful to the staff of the Turkana Basin Institute and the Mpala Research Centre for their support. We thank our research assistants affiliated with the Arizona State University team for aiding with data collection: E. Carlystus, A. Lotiira, C. Muya, G. Topos, and D. Lomelu. The authors acknowledge Research Computing at Arizona State University for providing high-performance computing resources that have contributed to

the research results. The National Museums of Kenya provided institutional support to conduct the research in Kenya. We thank W. Tong and J. Grenier for scientific discussions. We thank I. Wallace for thoughtful comments on a previous version of the manuscript. Above all else, we thank our participants and host communities for their hospitality, feedback, and for their continued support in this work. **Funding:** This work was funded by internal awards from Princeton University to J.F.A., The John Templeton Foundation (grant 48952 to S.M.), The National Institute of General Medical Sciences of the National Institutes of Health (NIH) (R35GM124827 to M.A.W. and R35GM147267 to A.J.L.), and The National Institute of Environmental Health Sciences of the NIH (F32ES034668 to K.M.G.). It was also supported, in part, by the Intramural Research Program of the National Human Genome Research Institute, National Institutes of Health. A.M.T.O. was supported by The Graduate College at Arizona State University, The Achievement Rewards for College Scientists Foundation Phoenix Chapter as a Pierson Scholar, and Arizona State University Sigma Xi. A.J.L. was also supported by the Helen Hay Whitney Foundation, a Searle Scholars Award from the Kinship Foundation, and an Azrieli Global Scholars Award from the Canadian Institute for Advanced Research. **Author contributions:** Conceptualization: A.J.L. and J.F.A. Data curation: A.J.L., J.E., K.M.G., J.K., and B.M. Formal analysis: A.J.L., I.V.C., E.R.G., C.H., P.K., S.L., T.P., R.S., and J.F.A. Funding acquisition: A.J.L., A.C.S., S.M., M.A.W., and J.F.A. Investigation: Field – A.J.L., K.M.G., C.H., S.M., R.D., C.M., J.K., B.M., J.P., J.D.R., A.R., and J.F.A.; Lab – A.J.L., J.E., J.K., F.L., D.M., C.M., R.M., B.M., and J.F.A. Methodology: A.J.L., I.V.C., and P.W.M. Project administration: A.J.L., S.N.N., D.M., and J.F.A. Resources: A.C.S., S.M., and M.A.W. Visualization: A.J.L. and I.V.C. Writing – original draft: A.J.L. and J.F.A. Writing – review & editing: A.J.L., I.V.C., K.M.G., E.R.G., M.G., C.H., J.K., S.N.N., R.D., R.M., D.M., P.W.M., P.K., S.L., T.P., R.S., A.C.S., S.M., and M.A.W. **Diversity, equity, ethics, and inclusion:** This work is the result of several long-term collaborations with Turkana communities as well as other groups in the region, built on mutual respect and shared investment in the research process. Our research teams include Kenyan scientists and students, and our study was shaped by years of conversation with local leaders, health workers, and participants. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** The genomic data generated in this study have been submitted to (i) the NIH database of Genotypes and Phenotypes (dbGaP) under accession number phs002219.v1.p1 and (ii) Zenodo (58–60). Analysis code is available at [https://github.com/AmandaJLea/Turkana\\_selection](https://github.com/AmandaJLea/Turkana_selection); [https://github.com/SexChrLab/Kenya\\_Selection\\_and\\_Demography/tree/master/Population\\_History](https://github.com/SexChrLab/Kenya_Selection_and_Demography/tree/master/Population_History); and <https://github.com/ayroles-lab/turkana-stc1>. The analysis code has also been deposited at Zenodo (58). Raw immunoblotting images and communication materials used to return results to study participants are also available at [https://github.com/AmandaJLea/Turkana\\_selection](https://github.com/AmandaJLea/Turkana_selection) as well as Zenodo (58). **License information:** Copyright © 2025 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. <https://www.science.org/about/science-licenses-journal-article-reuse>

## SUPPLEMENTARY MATERIALS

[science.org/doi/10.1126/science.adv2467](https://science.org/doi/10.1126/science.adv2467)

Materials and Methods; Figs. S1 to S24; Tables S1 to S22; References (61–126)

Submitted 11 December 2024; accepted 30 July 2025

10.1126/science.adv2467