A domestication history of dynamic adaptation and genomic deterioration in sorghum.

Oliver Smith[1,2], William V Nicholson[1,3], Logan Kistler[1,4], Emma Mace[5], Alan Clapham[1], Pamela Rose[6], Chris Stevens[7], Roselyn Ware[1], Siva Samavedam[1], Guy Barker[1], David Jordan[8], Dorian Q Fuller[7], Robin G Allaby[1*].

1. School of Life Sciences, University of Warwick, Coventry, CV4 7AL, United Kingdom.
2. Natural History Museum of Denmark, Øster Voldgade 5-7**,** 1350 København K, Denmark.
3. Warwick Medical School, University of Warwick, Coventry, CV4 7AL, United Kingdom.
4. Department of Anthropology, Smithsonian Institution, National Museum of Natural History, Washington, D.C. 20560, USA.
5. Department of Agriculture, Fisheries and Forestry Queensland (DAFFQ), Warwick, Queensland 4370, Australia.
6. The Austrian Archaeological Institute, Cairo Branch, Zamalek, Cairo, Egypt
7. Institute of Archaeology, UCL, London, United Kingdom
8. Queensland Alliance for Agriculture and Food Innovation, The University of Queensland, Warwick, Queensland 4370, Australia.

* Corresponding author

Abstract

The evolution of domesticated cereals was a complex interaction of shifting selection pressures and repeated episodes of introgression. Genomes of archaeological crops have the potential to reveal these dynamics without being obscured by recent breeding or introgression. We report a temporal series of archaeogenomes of the crop sorghum (*Sorghum bicolor*) from a single locality in Egyptian Nubia. These data indicate no evidence for the effects of a domestication bottleneck, but instead suggest a steady decline in genetic diversity over time coupled with an accumulating mutation load. Dynamic selection pressures acted sequentially to shape architectural and nutritional domestication traits, and to facilitate adaptation to the local environment. Later introgression between sorghum races allowed exchange of adaptive traits and achieved mutual genomic rescue through an ameliorated mutation load. These results reveal a model of domestication in which genomic adaptation and deterioration was not focused on the initial stages of domestication but occurred throughout the history of cultivation.

Keywords

Ancient DNA, archaeobotany, bottleneck, introgression, genomic rescue

The evolution of domesticated plant forms represents a major transition in human history that facilitated the rise of modern civilization. In recent years, our understanding of the domestication process has become revised considerably (1). In the case of cereals, it has been recognized that the selective forces that give rise to domestication syndrome traits such as the loss of seed shattering were generally weak—comparable to natural selection (2,3)—and that the intensity of selection pressures changed over the course of time as human technology evolved (4). Furthermore, domesticated lineages have often been subjected to repeated introgressions from local wild populations that endowed adaptive traits and obscured historical signals in the genome (5-7). Such complexity obfuscates attempts to reconstruct the evolutionary history of domesticated species from modern plants. To counter these confounding factors, in this study we directly tracked the evolutionary trajectory of a domesticated species, sorghum (*Sorghum bicolor* ssp. *bicolor* (L.) Moench.), through the archaeological record. This approach enabled the identification of selection pressures not clear today, and the tracking of the introgression process, revealing a domestication history which runs counter to the expectations of the current conventional model of domestication.

Sorghum is the world's fifth most important cereal crop and the most important crop of arid zones (8) used for food, animal feed, fibre, and fuel. The evolution of sorghum has seen it's transition from being a wild pluvial plant in north-eastern Africa (*S. bicolor* ssp. *verticilliflorum* (Steud.) De Wet ex Wiersema & Dahlberg, hereafter referred to as *S. verticilliflorum* for clarity) to the ancestral

domesticated form *Sorghum bicolor* type bicolor in Central Eastern Sudan by around 5000 years ago, while cultivation is inferred to have begun by 6000 years before present (yrs BP) (9). Ultimately, four specialized agroclimatic adapted types evolved after domestication—durra, caudatum, guinea, and kafir (10-12). The derived types were likely founded on introgressions of the wild progenitor complex *Sorghum verticilliflorum* or closely related species into the ancestral bicolor type, endowing traits such as drought tolerance in the case of type durra (10,13). The evolutionary history of sorghum, replete with introgression, is difficult to reconstruct from modern datasets. However, a temporal series of archaeobotanical domesticated sorghums spanning back to 2100 yrs BP at the archaeological site of Qasr Ibrim, situated on the Nubian frontier of northern Africa, affords the opportunity to track this complex crop directly through time removing the obscuring effects of introgression (14). Prior to this, apparently wild sorghum is present at Qasr Ibrim from at least ca. 2800 yrs BP. Domesticated sorghum (race bicolor) appears at the site ca. 2100 yrs BP. After this time period phenotypically domesticated sorghum of the ancestral type bicolor occurs throughout all cultural periods until the site was abandoned 200 years ago. During the early Christian period at 1470 yrs BP, the oldest known drought-adapted, free-threshing durra type appears at the site and occurs there for the rest of the site's occupancy. The origins of the durra type are unclear. Current distributions in northern and eastern Africa and its dominance in the Near East and South Asia led to the proposal that durra originated on the Indian subcontinent (10,15). It is now thought that durra arose after an African bicolor

type reached the subcontinent 3.8 Kyrs BP and hybridized with an indigenous wild species, with durra spreading to Africa at some point after 1800 yrs BP (13,16).

Results

Genetic diversity of sorghum over time

To gain a longitudinal insight into the evolutionary history of sorghum, we sequenced 9 archaeological genomes from different time points at Qasr Ibrim, including a wild phenotype from 1765 yrs BP and 8 domesticated phenotypes between 1805 and 450 yrs BP, a further 2 genomes from herbarium material, and 12 genomes of modern wild and cultivated sorghum types representing the varietal range (Table S1, S2). We investigated how genetic diversity has changed through time by measuring heterozygosity within genomes. First, we validated the use of genomic heterozygosity as a measure of genetic diversity, since heterozygosity can become depressed through excessive inbreeding, for example in the maintenance of germplasm . To do this, we determined the effective mating strategy of modern, wild, and ancient genomes based on their relative heterozygosity and comparable nucleotide diversity values (see methods), Figure S1, Table S3. Wild sorghum is expected to have a mating strategy of around 40% inbreeding while land races are expected to be closer to 80% inbreeding (17). One of our four wild genomes (race verticilliflorum) showed an appropriate level of effective mating strategy and so passed our validation, which we used as representative of sorghum wild progenitor diversity. The modern bicolor reference genome (BTX623) has a high effective mating strategy

of between 90-95% indicative of an elite inbred line. The ancient genomes showed intermediate mating strategies, suggesting that the switch from 40% to 80% inbreeding may have been a gradual process.

The heterozygosity of 100 kbp genomic blocks across the genome revealed a pattern of broad variation in heterozygosity in the wild progenitor *S. verticilliflorum* that became progressively narrower over time in the ancestral bicolor type, Figure S2. The remaining wild genomes all showed narrow and low ranges of heterozygosity indicative of inbreeding, as suggested by our validation, Figure S3.

The wild phenotype sorghum at Qasr Ibrim (sample A3) has a narrower variation in heterozygosity than the wild progenitor (represented by modern wild diversity), suggesting that it had been already been subject to genetic erosion, and it post-dates the earlier bicolor type sample A5. Together, this evidence suggests that sample A3 is likely a feral form of sorghum resulting from introgression between wild and cultivated populations. Conversely, the durra types all showed similar low levels of genomic variation in heterozygosity suggestive of genetic erosion prior to their appearance at Qasr Ibrim, Figure S2, S4. Total genomic heterozygosity of bicolor over time confirmed that the feral 'wild' ancient sorghum had already undergone considerable genetic erosion relative to the wild progenitor. To our surprise, the decreasing trend in heterozygosity over time fits a linear model (*p* values 0.0041 and 9.2×10^{-6} for parameters a and b respectively) better than an exponential model (*p* values 0.042 and 9.3×10^{-5}) as would be expected from an early initial loss of diversity

through a domestication bottleneck (18), Figure 1, Table S4 suggesting that there was no measurable effect on genetic diversity attributable to a domestication bottleneck.

We further examined the pairwise sequentially Markovian coalescent (PSMC) profile of the bicolor type genomes (19) to generate a profile of past population size from the genomic diversity, Figure S5. This produced an estimate of a long-term trend of decreasing population size over the past 100 Kyrs, a pattern similar to that previously observed in maize (20) and rice (21). This pattern is produced spuriously when sampling a single deme from a larger metapopulation that has received gene flow over time (22), which is likely to be the case for sorghum. All genomes also produce a strong signal of population expansion after 2000 years ago, which can also be produced by hybrid genomes between two populations (22,23). While there is no clear signal in these data for a domestication bottleneck, the profiles are indicative of long-term gene flow followed by a hybridization event between two populations that closely matches the arrival of the durra type sorghum at Qasr Ibrim.

Mutation load over time in Sorghum

The apparent lack of a domestication bottleneck runs contrary to expectations for a domesticated crop. To investigate the apparent lack of a domestication bottleneck further, we considered the mutation load. An expected consequence of the bottleneck is a rise in mutation load as small populations incorporate

deleterious mutations through strong-acting drift. High mutation loads have generally been observed in domesticated crops (20,24,25), which have been taken as a confirmation of the effects of the domestication bottleneck, although other causes are possible such as hitchhiking effects during selection. We measured the mutation load over time in the archaeological sorghum using a genome evolutionary rate profiling (GERP) analysis considering the total number of potentially deleterious alleles (26) (see methods), Figure 2. As with other domesticated crops, modern sorghum has a higher mutation load than its wild progenitor under both recessive and additive models (20). In contrast to the expectations of a domestication bottleneck we did not observe an initial large increase in mutation load associated with domestication, but rather an overall increasing trend in mutation load over time to the present day suggesting a process of load accumulation combined with selective purging episodes. In this case the trend line is well described by either a positive exponential model that approaches a straight line (*p* values 1.08×10^{-12} and 0.0683 for parameters *a* and *b* respectively) or a linear model (*p* values 1.07×10^{-12} and 0.0686), Table S4, suggesting mutation load may have become increasingly problematic in recent times. However, the *p* values suggest that the coefficient for the time in each model is only weakly significant, which could be the result of multiple processes ongoing, such as a strong increase in the rate of mutation load accumulation in recent times. When we considered the number of sites containing deleterious alleles (dominant model) rather than total number of deleterious alleles, we observe a decreasing trend over time (Figure S6). This pattern suggests that part

of the rising mutation load in the bicolor type was due to the increased homozygosity over time causing fixation of deleterious alleles originating from the wild progenitor pool. This notion is supported by the observation that 0.43-0.48 of genomic blocks associated with a decrease in the number of deleterious sites over time are associated with an increase in homozygous deleterious sites, Figure S7, Table S5. However, there is also evidence for expansions in deleterious sites in both sample A5 and BTX623 that are not associated with increased homozygosity indicating new influxes of mutation load not originating from the wild progenitor, Table S5. There is variation over time in mutation load, most notably in 1805 year-old sorghum (sample A5) that shows a sharp increase due to the incorporation of strongly deleterious alleles, both in the total number of alleles and number of sites. Interestingly, we found that the durra types show a pattern that contrasts to the bicolor type with relatively little change in heterozygosity and a significant fall in mutation load over time, suggesting the purging of deleterious mutations either through selection or genomic rescue through hybridization (Figure S4, S6). The contrasting patterns in mutation load over time are also reflected in methylation state profiles, which can reflect the state of genome-wide stress (27), Figure S8.

Signals of selection in sorghum

We considered that episodes of selection could have contributed in part to the variation in mutation load observed over time, either through reducing population size due to the substitution load or through hitchhiking effects. Three approaches

were used to identify candidate regions under selection. Firstly, we surveyed for wild/domestication heterozygosity to look for significant reduction in heterozygosity in domesticates, which revealed 30 peaks of genome-wide significance (denoted by prefix pk), Figure 3, Table S6, S7. We also specifically surveyed 38 known domestication loci and also found a significant reduction of heterozygosity in 15 of the 38 associated regions, Table S8. We further characterized the domestication loci in their pairwise nucleotide distance to the modern domestic alleles in BTX623, Figure S9. Secondly, we used a SweeD analysis to detect selective sweeps (28), which identified 11 peaks (denoted by prefix s), Figure 3, Table S9. In the third approach we utilized the temporal sequence of archaeogenomes to investigate episodes of selection intensification by considering the gradient of heterozygosity change over time (see methods). In this latter approach we tracked the gradient of change in the heterozygosity of regions identified in the first two approaches and assigned significance based on the gradient deviation from the genome average for each type. We considered multiple time sequences representing alternative possible routes through contemporaneous genomes over time within type bicolor and type durra respectively. This revealed a period of selection intensification associated with domestication loci prior to 1805 yrs BP, with oscillations in diversity discernible in the data after this time, Table S10, Figures S10-S12.

Together, the selection identification approaches exploit a range of different types of signature left by selection, and reveal a complex and dynamic history of selection over time summarized in Figure S13. We generally found more

evidence for selection in the bicolor type sorghum than the durra type. Despite its apparent wild phenotype, the wild sorghum (A3) at Qasr Ibrim from 1765 yrs BP shows evidence of selection at domestication loci concerned with architecture (*int1*, *tb1*), suggesting possible introgression with contemporaneous domesticated forms (represented by sample A5) that could have contributed to reduced heterozygosity. Interestingly, the intensification signals show some overlap between samples A3 and A5 (*int1* and *ae1*), with A5 showing further evidence for selection at shattering, dwarfing and sugar metabolism loci (*Sh3/Bt1*, *dw2* and *SPS5*) that would contribute to the domesticated phenotype of A5 relative to A3. Subsequent to this a period of intensification in selection is apparent both for dwarfing and sugar metabolism traits (710-715 years BP) in the bicolor type, with ten domestication loci showing significantly low levels of heterozygosity in this lineage by 710 yrs BP in A7. In the bicolor type, two of the sugar metabolism associated gene families show evidence of early selection controlling photosynthetic sucrose production first (*SPS*) and then an intensification of selection for breakdown (*SUS*). A third gene family (*SUT*) associated with sucrose transport appears to come under later selection in bicolor. In contrast, fewer domestication loci were found to show evidence of intensifying selection in the durra type, and none showed evidence of low heterozygosity. In this case we detected signals for an intensification of selection on tillering and maturity associated loci (*gt1*, *ma3*, the latter also being detected using SweeD s1 in the bicolor lineage). Significant heterozygosity reduction was identified in windows containing a large number of disease resistance loci (pk4,

pk11, pk15, pk20, pk24, pk25) as well as sugar metabolism loci (pk14, pk18, pk19, pk22) in the bicolor type. One of SweeD peak (s2) was closely matched to pk5 on chromosome 2 in the 54.0 – 54.2 Mbp interval, possibly indicating signatures for the same selection process. This region shows a consistently low heterozygosity over time in the bicolor type with the notable exception of the 1805 year old sorghum (A5). The region contains the far-red impaired response genes (*FAR1*), as well as anther indehiscence 1 (*AI1*). The *FAR1* gene is associated with phytochrome A signal transduction (29), so is important in responses to far red light that divert resources away from tall growth to increase root and grain growth. The *AI1* gene regulates anther development (30), allowing earlier development. Either of these genes may be locally adapted to the Qasr Ibrim environment since they already appear to be under intense selection in the wild sorghum at this site (sample A3), but not apparently under as much constraint in modern sorghum type bicolor.

The dynamic selection over time detected with most intensification of selection occurring before 1805 years BP, appears to correlate with a sharp increase in mutation load in the bicolor type. In contrast, the durra type shows much less evidence of selection and on arrival at Qasr Ibrim shows initially similar levels of mutation load to the bicolor type that then decreased over time (Figure S6). To investigate whether loci of selection are associated with higher regions of mutation load we measured the maximum deviations between genomes in GERP load scores across the genome and compared those to the locations of selection peak candidates (Figure 3, Table S10). Selection

signatures were highly significantly associated with regions of maximum deviation in mutation load with 30% of low heterozygosity peaks (p 8.04 x10^{-9}), 45% SweeD peaks (p 2.55x10^{-7}) and 26% domestication loci (p 5.03 x 10^{-5}) occurring in such regions. The intensification of selection is associated with increased mutation load and could explain the spike in mutation load observed in the 1805 year old sorghum (sample A5).

Genome rescue through hybridization

We considered that the decreasing mutation load observed in the durra type could be due to a genomic rescue caused by hybridization with the local bicolor type. To investigate for evidence of hybridization we first constructed a maximum likelihood phylogenetic tree of wild and cultivated total genomes (Figure S14), and individual trees for 970 sections across the genome (Supplementary data set 1). After accounting for biases introduced by ancient DNA modification (see methods), both the durra and bicolor type from Qasr Ibrim form a single clade to the exclusion of modern bicolor and durra types, suggesting they have indeed hybridized over time. D-statistic analysis (31) shows over time the durra type moved from a genomic background highly differentiated to that of bicolor and *S. verticilliflorum*, possibly relating to the putatively hybrid origin of durra (13,16), to becoming increasingly similar to the local bicolor type, suggesting progressive introgression between the two types (Figure S15). To estimate the proportion of the durra genome attributable to the hybrid origin background we examined fd statistics. These show that in the case of the oldest durra (A11) about 25% of the

genome is of the highly differentiated background, while in the younger A10 this proportion has been reduced, Figure S16. We then compared the archaeological genomes against a global sorghum diversity panel (32,33) (Figure 4, Supplementary video). The archaeological genomes are distributed along an axis of spread that has Asian durra types at the extremity. The oldest archaeological durra type (A11) sits between East African durra types and Asian durra types, whilst the wild phenotype sorghum, most closely aligned to the subsequent type bicolor, sits close to the center of the PCA, suggesting East African durras may have arisen from a hybridization between Asian durra and African bicolor. The oldest archaeological durra type in this study (sample A11) may represent one of the earliest of the east African durras. The younger archaeological genomes of the two types become progressively closer on the PCA supporting a process of ongoing hybridization between the two types over time.

Finally, we investigated whether the hybridization between the bicolor and durra types led to adaptive introgression or genomic rescue. Phylogenetic incongruence between the bicolor and durra type clades suggests that hybridization was frequent at loci under selection (Table S12). In agreement with previous studies (12) there is incongruence at the dwarfing *dw1* allele between durra and bicolor, with a single durra type sample sitting within the bicolor type clade in this region. Interestingly, seven of the nine sugar-metabolism associated loci potentially under selection in the bicolor type are also areas of incongruence with durra. However, in the case of *SPS5* in which we identified early intensification of selection in the bicolor type, no phylogenetic incongruence

occurred. Conversely, at the maturity locus *ma3* containing region, the durra type A11 that was identified as potentially under selection sits within the bicolor clade. The *FAR1/AI1* loci region, which appears to have been under strong selection in bicolor throughout, also is a region of incongruence with durra.

Assuming that the two types, bicolor and durra, had accrued mutation loads independently for the 2 Kyr (13) prior to the introduction of the durra type to Qasr Ibrim, then hybridization would have afforded the opportunity for genomic rescue between the two types. We therefore considered all ancestor/descendent pairs of genomes within the bicolor and durra type lineages in the context of a third potential donor genome, and scanned all sites for comparative GERP load scores under the additive model. We calculated firstly the difference in GERP load scores between the ancestor and potential donor to give an 'unbiased rescue value' that reflects a donor's potential to reduce mutation load across the entire genome, Figure 5a. We secondly assessed the donor's potential to effect mutation load reduction specifically at only those sites in which there had been a reduction in GERP load score between the ancestor and descendent to give an 'on target rescue value'. In most ancestor/descendent/donor combinations the reduction of mutation load in the descendent could be explained by a reduced load value in the donor, but often the unbiased exchange of load across the genome would be expected to result in an overall increase in mutation load. We considered whether if any of these introgressions occurred there was a bias towards genetic exchange that reduces mutation load. To do this we considered the expected proportion of instances in which both the descendent and donor

would both have a lower or higher load than the ancestor by chance relative to the observed values, Figure 5b, Table S13. In general, there was a highly significant threefold increase in the co-occurrence of lower load in the descendants and donors, but only a marginal increase in the co-occurrence of higher load. This is strong evidence that one or more of these introgression scenarios occurred, and that there was a bias in genetic exchange in regions of lower load in the donor. In the case of bicolor, the most significant scenario is A5/A6/A11 for ancestor/descendent/donor respectively, and in the case of durra A11/A9/A7. Both of these scenarios make chronological sense and indicate that a process of mutual rescue between durra and bicolor occurred.

Discussion

This study demonstrates that sorghum represents an alternative domestication history narrative in which the effects of a domestication bottleneck are not apparent, mutation load has accrued over time probably as a consequence of dynamic selection pressures rather than a domestication-associated collapse of diversity, and that genomic rescue from load occurred when two different agroclimatic types met.

The linear nature of the decreasing trend in diversity over time observed in sorghum in this study is surprising. An extreme bottleneck early in the history of would be expected to lead to a strong negative exponential trend as diversity is rapidly lost in the early stages of domestication. An alternative explanation for the trend could be that diversity has been lost steadily through drift over time.

However, a simple drift model shows that such a ten-fold loss in diversity would also be associated with a negative exponential trend, Figure S17. It is possible that diversity loss could have been supplemented by gains through introgression from the wild over time, counteracting the trend made by drift. Sample A3 could be the result of a wild introgression event since there are older domesticate phenotypes in the archaeobotanical record, such as sample A5. Sorghum is known for its extensive introgression leading to a strong regional structure within cultivars (12), making continuous introgression seem like a plausible scenario for sorghum at Qasr Ibrim. Incorporation of three systems of introgression into the simple drift model in which introgression is either constant, diminishing or increasing over time still results in a non-linear trend, which become parabolic when introgression becomes very high over time (not shown), Figure S13. We therefore think it unlikely that a model of constant drift and introgression is causative of the apparent linear decrease in diversity over time observed in this study.

The transition in mating system from 40% to 80% inbreeding over time in sorghum would be expected to reduce heterozygosity, and a model of constant change in mating system that is supported by the observations in this study would be expected to produce a linear decrease. However, we calculate that the ten-fold reduction in heterozygosity observed would require a 90-95% change in inbreeding, much wider than that observed in sorghum, making this unlikely to be a complete explanation.

Such linear decreases in diversity have been observed in human populations with increasing geographic distance from Africa and are most robustly explained by sequential founder models (34), and a similar observation has been made in maize (20). The annual cycle of crop sowing and harvesting in sorghum also represents a serial founding event scenario. Ethnobotanical evidence shows that only 1% of the sorghum population contributes to the next generation as farmers set aside a small number of plants to be the parents of the next generation, in contrast to practices with other cereals such as wheat and barley where seeds are sampled from across the whole population (35). We explored five population simulation models in which diversity was lost over time. Two models included a 1% annual founding event, one with a switch from 40% to 80% inbreeding in a single step at the point of domestication (model 1), the other with a gradual shift in mating systems from 40% to 80% inbreeding over 6000 years (model 2). The remaining three models lacked founding events, but respectively had either a gradual switch in mating system (model 3), a single step mating system switch at domestication (model 4), or no change in mating system and a continued 40% inbreeding level (model 5).

For each model we determined the population size associated with the gradient of heterozygosity loss observed in our data from the linear regression of gradients of loss associated with differing orders of magnitude of population size (Figure S18, Table S14). We then simulated the models at their respective population sizes over 6000 generations, Table S15. All models with the exception of model 2 had a better fit to an exponential rather than a linear trend. Model 2

was the closest fit to a linear descent in diversity (R^2 = 0.97), model 3 gave a weaker fit (R^2 = 0.90). The population size associated with model 3 was small (N=15955), while model 2 was considerably larger (N=727778). Given sorghum planting densities (35), model 3 equates to a 4-hectare system, while model 2 equates to 204 hectares. Subsistence farmers typically utilize a single one-hectare field of sorghum (35), reflecting highly localized and regional levels of diversity for models 3 and 2 respectively. Sorghum genetic diversity is structured on the regional scale in Africa within small language groups (36) in agreement with the regional scale represented in model 2. Together this evidence supports a scenario in which genetic diversity was lost in sorghum at a regional scale due to the founding effects of the agricultural practices involved. This may explain why similar large losses of diversity are not observed in other cereals such as maize and barley (37). The process likely incorporated all the available wild genetic diversity at the outset rather than a substantial initial domestication bottleneck.

The deleterious effects of mutation load are becoming increasingly apparent and a major problem in modern crops such as the dysregulation of expression in maize (38). The study here demonstrates the potential immediacy of the problem in that mutation load may generally be a consequence of recent selection pressures leading to an exponentially rising trend, rather than a legacy of the domestication process. While the general trend of the archaeogenomes is for the increase in the number of sites homozygous for deleterious variants (recessive model), the overall number of sites holding deleterious variants decreases

(dominant model). This trend suggests a general purging of variants from the standing variation of the wild progenitor combined with the rise of homozygosity with decreasing diversity of the variant sites that remain. However, this is sharply contrasted by modern sorghum in which there is a leap in the number of sites holding deleterious mutations (dominant model). This process contributes to the accompanying jump in load under both the recessive and additive models in modern sorghum. This indicates a large influx of new deleterious variants within the last century giving the trend of mutation load accumulation an exponential shape. It is likely that this influx of mutation load is the product of recent breeding programs and the genetic bottlenecks associated with the Green Revolution. The accumulation of load has previously been associated with mutation meltdown and extinction of past populations (39) but it remains unclear whether crops could follow the same fate in the absence of rescue processes, or whether such episodes could have been involved with previous agricultural collapses when crops experienced extensive adaptive challenges (40,41). In the case of sorghum, wild genetic resources may be valuable not only as a source of improved and environmentally adaptive traits, but also as a source for reparation of genome wide mutation load that may affect housekeeping and economic traits alike.

This represents the first plant archaeogenomic study that tracks multiple related genomes to gain insight into changes in diversity over time directly. The trends revealed, based on a relatively low number of archaeological genomes, suggest a domestication history contrary to that typically expected for a cereal

crop. Further archaeogenomes may establish whether this is a general trend for sorghum and other crops.

Methods

1. Sample Acquisition. Archaeological samples were sourced from A. Clapham from the archaeological site Qasr Ibrim, outlined in Table S1. For details on dating see section 1.3 below. Historical samples from the Snowden collection were sourced from Kew Gardens, Kew1: Tsang Wai Fak, collection no. 16366 Kew2: Tenayac, Mexico, collection assignation 's.n.'. Modern samples of *S. bicolor* ssp. *bicolor* type bicolor, durra, kafir, caudatum, drumondii and guinea were supplied through the USDA [accession numbers PI659985, PI562734, PI655976, PI509071, PI653734 and PI562938 respectively]. Wild sorghum samples *S. verticilliciliflorum, S. arundinaeum, and S. aethiopicum* were also obtained from the USDA [accession numbers PI520777, PI532564, PI535995], and wild *S. virgatum* was donated by D. Fuller. The outgroups *S. propinquum* and *S. halapense* were obtained from the USDA [accession numbers PI653737 and Grif 16307] respectively.

The genomes generated in this study were also compared to 1023 re-sequenced genomes taken from Thurber et al 2013 (33).

1.2 *A note on taxonomy.* The sorghum genus is complex with numerous taxonomic systems. After Morris *et al.*'s findings (12), we have elected not to describe the principal cultivar types as subspecies or races but rather simply

‘types’ to reflect the reality that there is evidence of considerable introgression between each of these forms. The wild progenitor of domesticated sorghum is a complex made up of four ‘races’ verticilliflorum, arundinaceum, aethiopicum and virgatum. However, the integrity of these races is also questioned, and the currently more accepted designation is one species, verticilliflorum, of which the other races are subtypes. For clarity and simplicity in this study we have used the race type as a variety designation. The subtype we used in the analysis as representative of the wild sorghum progenitor was race verticilliflorum for two reasons. Firstly, the remaining three genome types we sequenced all showed low levels of heterozygosity compatible with a high degree of germ plasm inbreeding leading to effective mating strategy estimates of 85% and 95% (see main text), inappropriate for wild sorghum which is expected to be closer to 40%. The second reason was that race types verticilliflorum and aethiopicum would be geographically more appropriate wild ancestors relative to the other two subraces being distributed across northern Africa.

1.3 A note on Qasr Ibrim and archaeological context of samples. Qasr Ibrim was a fortified hilltop site in the desert of Lower Nubia on the east bank of the Nile, about 200 km, south of Aswan in modern Egypt. It has been excavated over numerous field seasons, since 1963 by the Egyptian Exploration Society (UK). In recent years with higher Lake Nasser levels only upper parts of the site are preserved as an island (42,43). The desert conditions provided exceptional organic preservation by desiccation with exceptional preservation of a wide range

of biomolecules (e.g. 44-46). Systematic sampling for plant remains was initiated in 1984 (47) and the first studies of these remains were carried out in the 1980s by Rowley-Conwy (48) and had continued by Alan Clapham (49,50). The exceptional plant preservation has previously allowed successful ancient genomic studies of barley (44) and cotton (45).

Qasr Ibrim was founded sometime before 3000 years BP. It had occupations associated the Napatan kings (Egyptian Dynasty 25: 747-656 BC), possible Hellenistic and Roman Egypt (3rd century BC to 1st c. AD), the Meroitic Kingdom (1st century to 4th century AD), and local post-Meroitic (AD 350-550) and Nubian Christian Kingdoms (AD 550-1300). Earlier periods are associated temples to Egyptian and Meroitic deities. After Christianity was introduced the site had a Cathedral. Later Islamic occupations finished with use as an Ottoman fortress. The site was abandoned in AD 1812. The Sorghum material studied here comes from a range of different contexts from excavation seasons between 1984 and 2000. While the chronology of the site is well established by artefactual material, including texts in various scripts, several sorghum remains or associated crops, were submitted for direct AMS radiocarbon dating, as listed below in Table S2. For directly dated find the median of the 2-sigma calibrated age range has been used. Note that Radiocarbon calibration defines “the present” as AD 1950, and we have recalculated the median as before AD 2000, and assigned Snowden historical collections form the start of the 20th century as ca. 100 BP. For material not directly dated, sample A12 could be assigned based on associated pottery and finds, which have a well-established chronology through the Christian

periods (51), A12 is associated with Islamic/Ottoman material (1500-1800 AD, ca. 400 BP)

2. *DNA extraction.* DNA was extracted from archaeological and historical samples in a dedicated ancient DNA facility physically isolated from other laboratories. All standard clean-lab procedures for working with ancient DNA were followed. Single seeds from each accession were ground to powder using a pestle & mortar and incubated in CTAB buffer (2% CTAB, 1%PVP, 0.1M Tris-HCl pH 8, 20mM EDTA, 1.4M NaCl) for 5 days at 37°C. The supernatant was then extracted once with an equal volume of 24:1 chloroform:isoamyl alcohol. DNA was then purified using a Qiagen plant Mini Kit with the following modifications: a) 5x binding buffer was used instead of 1.5x and incubated at room temperature for 2 hours before proceeding. b) After washing with AW2, columns were washed once with acetone and air-dried in a fume hood to prevent excessive G-forces associated with centrifugal drying. c) DNA was eluted twice in a total of 100µl elution buffer and quantified using a Qubit high sensitivity assay.

DNA from modern samples was extracted using a CTAB precipitation method due to excessive polysaccharide levels precluding column-based extractions. Briefly, seeds were ground to powder and incubated at 60°C for 1 hour in 750 µl CTAB buffer as previously described, with the addition of 1µl β-mercaptoethanol. Debris was centrifuged down and the supernatant was extracted once with an equal volume of 24:1 chloroform:isoamyl alcohol. The supernatant was then collected and mixed with 2x volumes precipitation buffer (1% CTAB, 50mM Tris-

HCl, 20nM EDTA) and incubated at 4°C for 1 hour. DNA was precipitated at 6°C by centrifugation at 14,000 *g* for 15 minutes. The pellet was washed once with precipitation buffer and incubated at room temperature for 15 minutes before being centrifuged again under the same conditions. The pellet was dried and resuspended in 100µl high-salt TE buffer (10mM Tris-HCl, 1M NaCl) and incubated at 60°C for 30 minutes with 0.5µl RNase A. The DNA was then purified using Ampure XP SPRI beads.

3. *Library construction and genome sequencing.* Libraries for all samples were constructed using an Illumina TruSeq Nano kit, according to manufacturers' protocol. A uracil-intolerant polymerase (Phusion) was used to amplify the libraries, in order to eliminate the C to U deamination signal often observed in ancient DNA in favour of the 5' 5mC to T deamination signal. The purpose of this was to obtain epigenomic information after analysis using epiPaleomix (52). Consequently, the data set was reduced for non-methylated cytosine deamination signals in the 5' end, but showed expected levels of G to A mismatches for ancient DNA (5-10%) in the 3' end and high levels of endogenous DNA content typical for samples from this site (Table S1). We compensated for G to A mismatches and 5mC to T mismatches for later analyses, depending on potential biases that could be introduced (see sections 4 and 5). While this approach is thought to reduce library complexity by reducing the number of successfully amplified molecules, we considered this to be a worthwhile trade-off considering the exceptional preservation and endogenous

DNA content of the Qasr Ibrim samples. We found no evidence to suggest insufficient library complexity after amplification. A minor modification was made to the protocol for ancient and historical samples: a column-based cleanup after end repair was used, in order to retain small fragments that would otherwise be lost under SPRI purifications as per the standard protocol. Genomes were sequenced on the Illumina HiSeq 2500 platform. Ancient and historical samples were sequenced on one lane each using SR100 chemistry and modern samples on 0.5 lanes each using PE100 chemistry.

4. *Preliminary Bioinformatics processing.* Illumina adapters were trimmed using cutadapt v1.11 using 10% mismatch parameters. Resulting FastQ files were mapped to the BTX623 genome (53) using bowtie2 v2.2.9 (54) under --sensitive parameters. SAM files containing mapped reads with a minimum mapping score of 20 were then converted to BAM files using samtools v1.14 (55). To eliminate the possibility of downstream biases resulting from damage patterns typical of ancient DNA, deamination patterns were then identified and masked from BAM alignments using a Perl script of LK's design. The overall genome coverage of all archaeological samples was sufficient to ensure no low-coverage biases occurred. Variant calls format (VCF) files were then made from pileups constructed using samtools mpileup, and variant calls were made using bcftools v1.4 (55).

5. *Methylation analysis.* Since a uracil-intolerant polymerase was used for library generation, we analysed BAM files, without the masking treatment as described above, using epipaleomix (52) on the ancient samples. We then collated the number of identifiable 5mC sites globally for each sample. Epipaleomix is designed to characterise CpG islands typical to animal genomes and, is not suited to gene-specific analysis of plant genomes to due to their wider methylation states (CHH and CHG) (56). However, when assessing relative overall genome methylation between individuals of the same species, CpG islands measured in this way provide a perfectly adequate proxy. We opted for global and windowed-measurements to determine relative methylation states between samples.

6. *Evolutionary and population analyses.* Two archaeological genomes (A8 and A12) were from phenotypes intermediate between bicolor and durra types. We found that sample A8 was predominantly of bicolor type and A12 predominantly of durra type. Given the uncertainty of these samples and their likely hybrid origins, we elected to leave them out of most analyses.

6.1 *Heterozygosity analysis* The number of heterozygous sites was measured for each 100 kbp window of genome aligned to the BTX_623 reference sequence (53). The frequency distribution of heterozygosity was then calculated by binning the windows in 1 heterozygous base site intervals. Genomic heterozygosity can become depressed under conditions of inbreeding, such as through the

maintenance of germplasm. Therefore, we validated heterozygosity as a measure of genetic diversity by determining the extent to which inbreeding had reduced diversity relative to the population, and related this to the mating strategy. To do this we calculated the expected reduction in heterozygosity due to inbreeding from values of nucleotide diversity by determining the inbreeding coefficient F for a given mating strategy using the estimate *F* function of the *selection time* program (57). We then used the nucleotide diversity value for wild sorghum of 0.0038 (32) as a comparison to the genomic heterozygosity values obtained from the wild sorghum samples in this study. This yielded estimates of 40% inbreeding for our race verticilliflorum, which is close to the expected value for wild sorghum (17), Figure S1. The remaining wild sample genomes (aethiopicum, arundinaceum and virgatum) showed effective mating strategies of 85%, 95% and 95% respectively, indicative of extensive inbreeding. We therefore used the race verticilliflorum as our representative genome of wild sorghum. In the case of the ancient samples, we estimated a nucleotide diversity value by calculating the average pairwise distance between the A5, A6 and A7 genomes for bicolor types, and A11, A9 and A10 genomes for durra types, Table S3. Ancient bicolor showed an effective mating strategy of 60% inbreeding, while ancient durra was 50% inbreeding, Figure S1. In the case of A3, which was a feral type, there was no comparable nucleotide diversity to use for validation. Modern bicolor was compared to the elite line nucleotide estimate of Mace et al (32) of 0.0023, yielding an estimate of 90-95% inbreeding indicative of an inbred line.

Ratios of wild:cultivated heterozygosity were calculated for each window using *S. verticilliflorum* as the wild progenitor. Ratios closely approximate a negative exponential distribution. Probabilities of observed heterozygosity ratios for each window were obtained from a negative exponential distribution with λ equal to $1/\mu$ for all ratios for each chromosome. A Bonferroni correction was applied by multiplying probability values by the number of windows on a chromosome in Figure 4. Locations of 38 known domestication syndrome loci (shown in Tables S8 and S10) were obtained by reference to the BTX_623 genome. Candidate domestication loci were obtained from the scans of Mace *et al* (32). In the genome-wide scan peaks were considered significant if $1/p > 100$ after Bonferroni correction.

We considered the possibility that the observed heterozygosity levels may be influenced by postmortem DNA damage. To explore this, we characterized the relationships between time, heterozygosity and postmortem deamination. As we previously described, C to U damage signals are eliminated at the 5' ends of sequence reads because of our choice of polymerase, so we therefore characterized damage profiles at the 3' ends only, using mapDamage output statistic '3pGtoA_freq' and taking a mean of the 25 reported positions for each ancient or historical sample. Unsurprisingly, we found that the accumulation of damage patterns is a function of time in a logistic growth model, assuming a zero-point intercept for both factors ($R^2 = 0.9$). 80% of damage capacity under this model is reached reasonable quickly, in 331.0 years. All the Qasr Ibrim samples are at least 400 years old, and so we re-fitted a linear regression model

to these samples only so characterize these relationships in a true time-series. We found a negligible correlation between time and damage accumulation after 400 years (R^2 = 0.15, p = 0.34). Next, we characterized the relationship between age and heterozygosity under the same model (although without the assumption of a zero-point intercept, since even modern domesticate lines in this study show non-zero levels) and found a weak fit (R^2 = 0.64, p = 0.14). This relationship is however likely influenced by our central hypothesis, with 'less' domesticated samples being earlier in the archaeological record, and so a counter-argument should not be inferred from this analysis. Finally, we assessed the relationship between damage and heterozygosity by linear regression, assuming inappropriateness of a logistic model since both damage and heterozygosity factors are functions of time. We found a weak correlation when considering all samples (R^2 = 0.2, p = 0.2), and virtually no correlation when considering the Qasr Ibrim time series only (R^2 = 0.04, p = 0.61). Considering that the two historical Kew samples are ostensibly domesticates, and historical and geographic outliers to the rest of the dataset, we conclude that the observed levels of heterozygosity in the ancient samples are not influenced by postmortem damage patterns.

6.2 PSMC estimates

A Pairwise Sequential Markovian Coalescent (PSMC) approach was taken to infer population history for all ancient and historical samples, using the PSMC package (19). Input parameters were defined by the following: average coverage

depth per sample was calculated using the samtools depth function, and applied to –d (at one third) and –D (at 2x) parameters of the vcf2fq function of vcfutils.pl to generate diploid consensus sequences with bins of 100bp from soft-masked BAM files. The PSMC utility fq2psmcfa was then applied with a minimum mapping quality of 20. The PSMC algorithm itself was then run using the default parameters. The PSMC output was then rescaled and plotted using psmc_plot.pl, assuming a generation time of 1 year for an annual plant such as sorghum, and mutation rate of $3.0x10^{-8}$ substitutions per site per generation, being the closest estimate of mutation rates available from domesticates of the Panicoideae subfamily (58).

6.2 *Differential Temporal Heterozygosity Gradient Analysis.* Our rationale was to utilize the temporal sequence of genomes to identify time intervals associated with intensification of selection. To this end we designed an analysis to identify outliers in changing heterozygosity over time to the general genomic trend. We considered all possible historical paths between genomes given three pairs of samples were almost contemporaneous (A3/A5, A6/A7 and A9/A10), with wild *S. verticilliflorum* representative of the wild progenitor in the case of the bicolor lineages.

For each 100kbp window we calculated the gradient of change in heterozygosity between temporally sequential genome pairs by subtracting younger heterozygosity values from older and dividing through by the time interval between samples. Genome-wide gradient values for all 100kbp windows

were used to construct a non-parametric distribution to obtain probability values of change over each time interval for a 100kbp window between a particular pair of samples. Peak regions identified by heterozygosity ratio, SweeD analysis and known domestication syndrome genes were then measured for gradient probability.

6.3 *SweeD analysis.* VCF files from our 23 ancient, historical and modern samples and also 9 samples from Mace et al (32) were combined using the GATK (59) program CombineVariants. Subsequently, the combined VCF file was filtered - using bcftools v1.4 (55) - to only include sites with 2 or more distinct alleles and at sites where samples have depth less than 5 or a variant calling quality score less than 20 to exclude those samples. Then a further filter was applied - using bcftools v1.4 - to exclude variant calls due to C->T and G->A transitions relative to the reference, which potentially represent post-mortem deamination which has a high rate in aDNA samples (60). SweeD (28) was run with options for multi-threading (to run with 64 threads) and to compute the likelihood on a grid with 500 positions for each chromosome.

6.4 *Genome Evolutionary Rate Profiling (GERP) analysis.* This analysis was carried out broadly following the methodology of Cooper *et al.* (26). We aligned the repeat-masked genomes of 27 plant taxa to the BTX_623 sorghum reference genome using last, and processed resulting maf files to form netted pairwise alignment fastas using kentUtils modules maf-convert, axtChain, chainPreNet,

chainNet, netToAxt, axtToMaf, mafSplit, and maf2fasta. We forced all alignments into the frame of the sorghum reference using an expedient perl script, and built a 27-way fasta alignment excluding sorghum for GERP estimation. We created a fasta file of fourfold degenerate sites from chromosome 1 (347394 sites; NC_012870) with a perl script, and calculated a neutral rate model using phyloFit, assuming the HKY85 substitution model and the following tree:

(((((((((Trifolium_pratense,Medicago_truncatula),Glycine_max),Prunus_persica),(Populus_trichocarpa,Manihot_esculenta)),(((Arabidopsis_thaliana,Arabidopsis_lyrata),(Brassica_napus,Brassica_rapa)),Theobroma_cacao)),Vitis_vinifera),((Solanum_tuberosum,Solanum_lycopersicum),(Chenopodium_quinoa,Beta_vulgaris))),(((Zea_mays,Setaria_italica),(((Oryza_rufipogon,Oryza_longistaminata),Leersia_perrieri),(((Triticum_urartu,Aegilops_tauschii),Hordeum_vulgare),Brachypodium_distachyon))),Musa_acuminata)),Amborella_trichopoda)

We then calculated GERP rejected subsitutions (RS) scores using gerpcol with the default minimum three taxa represented for estimation. The mutation load for each genome was then assessed by scanning through their VCF files generated by alignment to BTX_623. Maize was used as an outgroup to judge the ancestral state, and only sites at which there was information from maize were incorporated into the analysis. Sites which differed to the ancestral state were scored based on the associated RS score for that site following the scheme of Wang *et al.* (20): 0, neutral, 0-2 slightly deleterious, 2-4, moderately deleterious,

>4 seriously deleterious. We collected scores under three models, recessive, additive and dominant. Under the dominant model we counted each site once regardless of whether it had one or two alternative bases to the ancestor, so giving the total number of base sites containing at least one potentially deleterious allele. Under the additive model we counted the total number of alleles that were alternative to the ancestor such that each homozygous alternative site scored 2, but heterozygous sites scored 1. Under the recessive model only sites that were homozygous for potentially deleterious variants were counted.

To investigate the significance of overlap between regions significant GERP regions of difference (GROD) between taxa and signatures of selection we used a binomial test in which the null probability of selecting a GROD was equal to the total number of GRODS (193) divided by the total number of 100 kbp regions studied (6598), and N and *x* were the total number of selection signals and the number of selection signals occurring in a GROD respectively.

We used the GERP profiles to explore potential genomic rescue from mutation load accrued independently in the bicolor and durra lineages prior to hybridization between the two types. For the purposes of this analysis we used the wild sorghum genome A3 as a possible wild ancestor genome to the domesticated bicolor form A5 even though this wild sample is contemporaneous to that domesticated form. All possible ancestor descendent pairs were assembled within bicolor or durra types, and all 100 kbp windows were scanned for the relative additive model GERP load scores for ancestor, descendent and a

third potential donor genome. The total potential for the donor genome to rescue the ancestral genome was scored summing the difference in GERP scores across all windows between the ancestor and donor. To better fit a scenario in which the donor genome was the causative agent of reduction GERP load score we identified windows that satisfied the condition ancestor GERP load score > descendent GERP load score, and summed up the difference in ancestor and potential donor scores to give an 'on target rescue' value (p1). We assessed whether the instances of on target rescue occurred significantly more frequently than would be expected by chance given the descendent and potential donor GERP profiles, by comparing the observed proportion of windows satisfying on target rescue with the expect proportion given the frequency of the component conditions of descendent < ancestor and donor < ancestor (e1). The number of windows satisfying a set of criterion out of a set number of windows follows a Binomial distribution, so we used the posterior Beta distribution to derive *p* values which is a better fit to the data than the Chi squared distribution in this case. Beta distributions were generated for both the observed and expected number of events, and the *p* values taken from the overlap of the two distributions representing the probability that both values could have been drawn from the same distribution. We repeated this analysis to compare the incidences in which both the descendent and the potential donor exceeded the GERP load score of the ancestor (p2), and their corresponding expected values (e2).

6.5 *Phylogenetics* Maximum likelihood tress were constructed using exaML (61) firstly using whole genome sequences (Figure S14), and for 970 consecutive blocks across the genome (supplementary data set). Prior to computing phylogenetic trees, the VCF files were processed as described in section 6.3 (on the SweeD analysis) albeit with our 23 ancient, historical and modern samples only.

The maximum likelihood tree using the whole genome sequences was constructed as follows. Our own script created a multiple sequence alignment file by concatenating the variant calls in the VCF file and outputting the results in PHYLIP (62) format. The program parse-examl from the ExaML package (version 3.0.15) was run in order to convert the PHYLIP format file into ExaML's own binary format. Also, ExaML requires an initial starting tree which was obtained by running (on multiple threads) Parsimonator v1.0.2, a program available as part of the RaxML package (63) - developed by the same research group - for computing maximum parsimony trees. An ExaML executable (compiled to run using MPI) was run on multiple CPUs in order to compute the maximum likelihood tree.

The trees for 970 consecutive blocks across the genome were computed by essentially the same approach as described above for a single tree, after a script obtained the blocks from the input VCF file (for the combined samples) and output them in PHYLIP format.

To assess potential donation between genomes at candidate loci we examined trees spanning the corresponding 100kbp windows. The tree topology

was examined for congruence in the maintenance of bicolor and durra type groups within the Qasr Ibrim group of genomes. Instances of phylogenetic incongruence were interpreted as candidate regions of recombination between the two genome types, although identification of the donor and recipient genomes was not possible using this approach. Simple cases in which a single genome from one sorghum type was found within the group of the other type were identified as incongruent taxa. In the case of regions that scored highly in the SweeD analysis no phylogenetic congruence was attempted because the taxon in which selection has operated is not identified.

6.6 *Principal Component Analysis of global diversity set.* A subset of 1894 common SNPs were identified between two separate data sets; the resequencing data of the 23 samples and GBS data from an unpublished set of 1046 diverse sorghum lines, which span the racial and geographic diversity of the primary gene pool of cultivated sorghum. The subset of SNPs were selected based on a minor allele frequency of >0.05 and frequency of missing data points <0.2. 580 of these diverse lines were described in Thurber et al (26). These lines were produced within the Sorghum Conversion Program which introgressed key height and phenology genes into exotic lines to enable them to be produced in sub-tropical environments. The introgressed regions spanned approximately 10% of the genome which were masked for the purposes of this analysis. Principal component analysis of the centered data matrix was performed in R (R core team, 2017) using the *prcomp* function in the base “stats” package.

6.7 D statistics. Patterson's D-Statistic and modified F-statistic on Genome wide SNP data was used to infer patterns of introgression (31). D-statistic and fd-statistic for each of the 10 chromosomes was calculated using the R-package PopGenome. Variant Call Format (VCF) file, which is generated after mapping reads of an individual sample to the reference genome, was given as input to the readVCF() function of the package (64).

We used four R-language based S4 class methods from PopGenome package to carry out the introgression tests for every chromosome. First, we used the method set.population by providing 3 populations (2 sister taxa and an archaic group) viz., P1=BTX_623, P2=varying samples, P3=*S. verticilliflorum*. Second, using set.outgroup function, we set an outgroup (P4= S.halapense). Third, the method introgression.stats was employed to calculate the introgession tests. Finally, we used jack.knife.transform method (64) which transforms an existing object belonging to GENOME class into another object of the GENOME class with regions that corresponding to a Jackknife window. Standard error was then calculated by eliminating one such window i.e., a single chromosome under study and calculation was applied to the union of all the other chromosomes.

We tested for admixture from the modern wild *S. verticilliflorum*, assuming this represents a genome prior to the appearance of the durra type on the African continent. The BTX_623 sorghum reference genome was taken as P_1, *S. verticilliflorum* was taken as P_3 and *S. halapense* was taken as the out group P_4. *S. halapense* is native to southern Eurasia to east India and does not readily

cross with *S. bicolor.* Samples were then tested at the P_2 position across all 100kbp windows, each chromosome tested separately. Negative values (indicating an excess of P_1/P_3 combinations) are expected when the BTX_623 genome is more similar to *S. verticilliflorum* than P_2. This is observed as expected for the durra types since they are thought to have genomic introgression from wild sorghum species on the South Asian subcontinent which is expected to unite P1 and P3 in regions where durra is different to both bicolor and its wild ancestor. The value of D decreases over time consistent with either an increase in instances of P_2/P_3 or instances of P_1/P_2, both suggesting progressive introgression between the durra and bicolor types over time. We affirmed this scenario by placing the oldest and youngest durra samples (A11 and A10 respectively) as P1. While A11 consistently shows highly positive values, indicating P2/P3 combinations, A10 shows a reduced positive D score with bicolors, and highly negative D scores with A11, indicating introgression from the verticilliflorum background, most likely directly from bicolors.

Positive values (indicating an excess of P_2/P_3 combinations) suggest a close relationship between *S.verticilliflorum* and P_2, which is observed the Qasr Ibrim bicolor types (A5, A6 and A7). This is expected if these ancient bicolors share diversity with the wild progenitor that has been subsequently lost in modern *S.bicolor.*

Since the fd statistic cannot be meaningfully interpreted from negative D statistic values, to investigate the extent to which the durra genomes are made up of an unknown genome putatively of South Asian origin we repeated the D

statistic analysis using the durra samples A10 and A11 at position P1 and examined the resulting fd statistics (Figure S16). Bicolor types group with verticilliflorum to the exclusion of durra (P2/P3 combinations) are in excess by 25%. These values lead us to conclude that about 0.25 of the durra genome is constituted of the unknown genome.

6.8 Linear and exponential line fitting to heterozygosity, GERP score data and simuation outputs.

A straight line was fit to the heterozygosity data in Figure 2 using the glm function (for generalized linear models) in R and also an exponential function was fit to the same data using the gnm package (for generalized non-linear models) in R obtaining the values for the parameters, standard errors, p and AIC shown in Table S3. (It was confirmed similar values were obtained for the parameters, standard errors, p and AIC by fitting the straight line model using the gnm package in R.)

6.9 Basic simulation of diversity loss through drift, introgression and serial founding events

To explore the effect on general trend line shape of introgression over time we used a basic simulation of drift loss using the standard equation:

$$\frac{H_t}{H_0} = \left(1 - \frac{1}{2N_{fo}}\right)\left(1 - \frac{1}{2N(\frac{N_e}{N})}\right)^{t-1}$$

where N_{fo}, N and N_e are the founding population size, census population size and effective population size respectively. For simplicity, we assumed in the case of our crop that all three population sizes were equal. To incorporate introgression we used a simulation to calculate and modify each generation by using the above equation to modify the diversity from the previous generation, and then adding a diversity value representative of gene flow. Gene flow was altered each generation by a power factor *f*, which was 1 in the case of constant introgression, 1.0001 in the case of diminishing introgression over time and 0.99995 in the case of increasing introgression over time, with an initial value for introgression as 0.000015, equating to the value of genetic diversity added to the population each generation. We used a founder population of 2000 for 6000 generations to recapitulate the observed 10-fold loss of diversity over this time frame in sorghum.

The serial founder event simulation was executed using the program *founderv6.pl* available for download at : (https://warwick.ac.uk/fac/sci/lifesci/(https://warwick.ac.uk/fac/sci/lifesci/research/archaeobotany/downloads/founderv6) following model:
To initialize, allele frequencies were randomly assigned for a defined number of alleles, and the associated heterozygosity calculated for a given breeding system for which the inbreeding coefficient F was calculated using the estimate F function of the *selection time* program (57). Frequencies were then adjusted randomly in a Markov Chain until a user defined starting heterozygosity was reached. In this case all simulations were initiated at a starting heterozygosity

equivalent to 0.003 and 40% inbreeding to reflect wild sorghum. The genotype frequency distribution was then calculated for the given inbreeding coefficient F and allele frequencies. Individuals were drawn from the population by randomly selecting two parent genotypes, gametes were either drawn from one or both parents with a probability determined by the mating system. Gametes were mutated to generate new alleles based on a general plant rate of 5×10^{-9} substitutions/site/year (65). N individuals were randomly drawn in this way, and the resultant allele frequency distribution calculated. A founder event was then generated by drawing Nb individuals from the allele frequency distribution, following the same methodology as above. The process was repeated for a defined number of cycles, with the inbreeding coefficient being recalculated each generation in simulations where the mating system changed over time.

We explored five model scenarios using this simulation, over several orders of magnitude of N (100, 1000, 10000, 100000, 1000000), for 1000 founding events, equating to 1000 years of agriculture. Each trial was repeated 100 times, equating to 100 genes being simulated independently. Model 1: 80% inbreeding population from the point of domestication with a 1% bottleneck each generation; model 2: dynamic mating system switch from 40% inbreeding which would reach 80% inbreeding over 6000 years with a 1% bottleneck each generation; model 3 dynamic mating system switch from 40% inbreeding which would reach 80% inbreeding over 6000 years with no bottlenecks; model 4 80% inbreeding population from the point of domestication with no bottlenecks; model 5 static 40% inbreeding with no bottlenecks.

The log of the initial gradient of loss of diversity plotted against the log of the population size for each model gave a linear relationship from which a linear regression was derived. The linear regression was then used to derive the population size for each model that corresponded to the observed gradient of diversity loss observed in the real archaeogenomic data, Table S14. Simulations were then carried out under the five different models and their respective predicted population sizes, each repeated ten times.

Data availability

Sequence data were deposited in the European Molecular Biology Laboratory European Bioinformatics Institute [project code PRJEB24962.].

Code availability

The serial founder event simulation was executed using the program *founderv6.pl* available for download at : (https://warwick.ac.uk/fac/sci/lifesci/(https://warwick.ac.uk/fac/sci/lifesci/research/archaeobotany/downloads/founderv6).

Acknowledgements

The authors would like to thank M. Nesbitt for permitting the use of herbaria material from Kew. OS, WN, GB and RGA were supported by the NERC (NE/L006847/1) and LK was supported by NERC (NE/L012030/1). CJS and DQF work with archaeobotanical materials was supported by a European

Research Council grant (no. 323842).

Competing interests

The authors declare no competing interests.

(1) Larson G, Piperno D, Allaby RG et al. Current perspectives and the future of domestication studies. *Proceedings of the National Academy of Sciences. U.S.A.* **111**, 6139-6146 (2014).

(2) Purugganan MD, Fuller DQ. The nature of selection during plant domestication. *Nature* **457**, 843-848 (2009).

(3) Fuller DQ, Denham T, Arroyo-Kalin M, Lucas L, Stephens C, Qin L, Allaby RG, Purugganan MD Convergent evolution and parallelism in plant domestication revealed by an expanding archaeological record. *Proceedings of the National Academy of Sciences U.S.A.* **111**, 6147-6152 (2014).

(4) Allaby RG, Stevens S, Lucas L, Maeda O, Fuller DQ. Geographic mosaics and changing rates of cereal domestication. *Philosophical Transactions of the Royal Society B*. **372**, 20160429 (2017).

(5) Poets AM, Fang Z, Clegg MT, Morell PL. Barley landraces are characterized by geographically heterogeneous genomic origins. *Genome Biol*ogy **16**, 173 (2015).

(6) Hardigan MA, Laimbeer FPE, Newton L, Crisovan E, Hamilton JP, Vaillancourt B, Wiegert-Rininger K, Wood JC, Douches DS, Farré EM et al. 2017. Genome diversity of tuber-bearing solanum uncovers complex evolutionary history and targets of domestication in the cultivated potato. *Proceedings of the National Academy of Sciences*, 114: E9999–E10008.

(7) Hufford M, Lubinksy P, Pyhäjärvi T, Devengenzo M, Ellstrand N, Ross-Ibarra J. 2013. The genomic signature of crop-wild introgression in maize. *PLoS Genetics*, 9: e1003477

(8) Food and Agriculture Organization of the United Nations. [http://www.fao.org/index_en.htm].

(9) Winchell F, Stevens CJ, Murphy C, Champion L, Fuller DQ. Evidence for Sorghum Domestication in Fourth Millennium BC Eastern Sudan: Spikelet Morphology from Ceramic Impressions of the Butana Group. *Current Anthropology* **58**, 673-683 (2017).

(10) Doggett H *Sorghum* 2nd Ed, Longman, Harlow (1988).

(11) Brown PJ, Myles S, Kresowich S. Genetic support for a phenotype-based racial classification in sorghum. *Crop Sci.* **51**, 224-230 (2011).

(12) Morris G, *et al.* Population genomic and genome-wide association studies of agroclimatic traits in sorghum. *Proceedings of the National Academy of Sciences U.S.A.* **110**, 453-458 (2013).

(13) Fuller, Dorian Q and Chris J. Stevens (n.d.) Sorghum Domestication and Diversification: A current archaeobotanical perspective. In: Anna Maria Mercuri, A. Catherine D'Andrea, Rita Fornaciari, Alexa Höhn (eds.) *Plants and People in Africa's Past. Progress in African Archaeobotany.* Springer (2017).

(14) Clapham AJ, Rowley-Conwy PA In Fields of Change–Progress in African Archaeobotany, Cappers R, ed. Groningen Archaeological Studies. Groningen, **5**, 157–164 (2007).

(15) de Wet JML, Harlan JR, Price EG Variability in *Sorghum bicolor*. In: Harlan JR, de Wet JMJ, Stemler ABL (eds) Origins of African plant domestication. Mouton Press, The Hague, p 453-463 (1976).

(16) Harlan JR, Stemler ABL The races of Sorghum in Africa. In: Harlan JR, de Wet J, Stemler ABL (eds) Origins of African plant domestication. Mouton Press, The Hague, p 465-478 (1976).

(17) Ohadi S, Hodnett G, Rooney W, Bagavathiannan, M. Gene Flow and its consequences in Sorghum spp. *Crit RevCritical Reviews in Plant SciSciences* 36:367-385 (2017).

(18) Meyer R, Purugganan M. Evolution of crop species: genetics of domestication and diversification. *Nature Reviews Genetics* **14**, 840-852 (2013).

(19) Li H, Durbin R. Inference of human population history from individual whole-genome sequences. *Nature* **475**, 493–496 (2011).

(20) Wang L, Beissinger TM, Lorant A, Ross-Ibarra C, Ross-Ibarra J, Hufford MB The interplay of demography and selection during maize domestication and expansion. *Genome Biol.* **18**, 215 (2017).

(21) Meyer RS *et al.* (2016) Domestication history and geographical adaptation inferred from a SNP map of African rice. *Nature Genetics* 48:1083-1088.

(22) Mazet O, Rodríguez W, Grusea S, Boitard S, Chikhi L 2016. On the importance of being structured: instantaneous coalescence rates and human evolution-- lessons for ancestral population size inference. *Heredity* 116: 362-371 (2016).

(23) Orozco-terWengel P. The devil is in the details: the effect of population structure on demographic inference. *Heredity* 116: 349-350. (2016).

(24) Renault S, Rieseberg L. The accumulation of deleterious mutations as a consequence of domestication and improvement in sunflower and other Compositae crops. *Mol. Biol. Evol.* **32**(9), 2273–2283 (2015).

(25) Liu Q, Zhou Y, Morrell PL, Gaut BS Deleterious variants in Asian rice and the potential cost of domestication. *Mol. Biol. Evol.* **34**(4), 908–924 (2017).

(26) Cooper GM *et al*. Distribution and intensity of constraint in mammalian genomic sequence. *Genome Res.* **15**, 901-913 (2005).

(27) Smith O, Clapham A, Rose P, Liu Y, Wang J, Allaby RG. Genomic methylation patterns in archaeological barley show de-methylation as a time-dependent diagenetic process. *Scientific Reports* **4**, 5559 (2014).

(28) Pavlidis P, Živkovic D, Stamatakis A, Alachiotis N SweeD: Likelihood-based detection of selective sweeps in thousands of genomes. *Mol Biol Evol* **30**, 2224–2234 (2013).

(29) Hudson M, Ringli C, Boylan MT, Quail PH The *FAR1* locus encodes a novel nuclear protein specific to phytochrome A signaling. *Genes Devel.* **13**, 2017-2027 (1999).

(30) Zhu QH, Ramm K, Shivakkumar R, Dennis ES, Upadhyahya NM The *ANTHER INDEHISCENCE1* gene encoding a single MYB domain protein is involved in anther development in rice. *Plant Physiol.* **135**, 1514-1525 (2004).

(31) Martin, S, Davey, J, Jiggins, C. Evaluating the Use of ABBA–BABA Statistics to Locate Introgressed Loci. *Mol. Biol. Evol.* **32**(1), 244-257 (2014).

(32) Mace ES *et al.* Whole genome sequencing reveals untapped genetic potential in Africa's indigenous cereal crop sorghum. *Nature Communications* **4**, 2320 (2013).

(33) Thurber CS, Ma JM, Higgins RH, Brown PJ Retrospective genomic analysis of sorghum adaptation to temperate-zone grain production. *Genome Biol.* **14**, R68 (2013).

(34) DeGiorgio M, Jakobsson M, Rosenberg N. Explaining worldwide patterns of human genetic variation using a coalescent-based serial founder model of migration outward from Africa. *Proc. Natl. Acad. Sci. U.S.A.* 106: 16057–16062 (2009).

(35) Alvarez N, Garine E, Khasah C, Dounias E, Hossaert-McKey M, McKey D). Farmers' practices, metapopulation dynamics, and conservation of agricultural biodiversity on-farm: a case study of sorghum among the Duupa in sub-sahelian Cameroon. *Biological Conservation* 121:533-543. (2005).

(36) Westengen OT, Okongo MA, Onek L, Berg T, Upadhyaya H, Birkeland S, Khalsa SDK, Ring KH, Stenseth NC, Brysting AK. Ethnolinguistic structuring of sorghum genetic diversity in Africa and the role of local seed systems. *Proc. Natl. Acad. Sci. USA* 111:14100-14105. (2014).

(37) Allaby RG, Ware R, Kistler L. A re-evaluation of the domestication bottleneck from archaeogenomic evidence. *Evolutionary Applications* doi.org/10.1111/eva.12680 (2018).

(38) Kremling KA *et al.* Dysregulation of expression correlates with rare-allele burden and fitness loss in maize. *Nature* 555:520-523 (2018).

(39) Rogers & Slatkin Excess defects in a woolly mammoth on Wrangel Island *PloS Genetics* 13(3): e1006601 (2017).

(40) Shennan S, Downey SS, Timpson A, et al. Regional population collapse followed initial agricultural booms in mid –Holocene Europe. *Nat Commun*. 4:2486 (2013).

(41) Allaby RG, Kitchen JL, Fuller DQ. Surprisingly low limits of selection in plant domestication. *Evolutionary Bioinformatics* 11:(S2) 41-51 (2016).

(42) Alexander, J The Saharan divide in the Nile Valley: the evidence from Qasr Ibrim. *The African Archaeological Review* **6**, 73-90 (1988).

(43) Rose P. Qasr Ibrim In: Bagnall RS, Brodersen K, Champion CB, Erskine A, Huebner SR (eds.) *The Encyclopedia of Ancient History.* Oxford: Wiley. Pp. 5695-5697 DOI: 10.1002/9781444338386.wbeah15340 (2013).

(44) Palmer SA, Moore JD, Clapham AJ, Rose P, Allaby RG. Archaeogenetic evidence of ancient Nubian barley evolution from six to two-row indicates local adaptation. *PLoS One*, ***4***(7), e6301 (2009).

(45) Palmer SA, Clapham AJ, Rose P, Freitas F, Owen BD, Beresford-Jones D, Moore JD, Kitchen JL, Allaby RG. Archaeogenomic evidence of punctuated

genome evolution in Gossypium. *Molecular biology and evolution*, ***29***(8), 2031-2038 (2012).

(46) O'Donoghue K, Clapham A, Evershed RP, Brown TA. Remarkable preservation of biomolecules in ancient radish seeds. Proc R Soc B Biol Sci. **263**, 541–547 (1996).

(47) Alexander, J , Driskell B. Qasr Ibrim 1984. *Journal of Egyptian Archaeology* **71**, 12-26 (1985).

(48) Rowley-Conwy P Nubia AD 0-550 and the 'Islamic' agricultural revolution: Preliminary botanical evidence from Qasr Ibrim, Egyptian Nubia. *Archeologie du Nil Moyen* **3**, 131-138 (1989).

(49) Clapham AJ, Rowley-Conwy PA. New discoveries at Qasr Ibrim, Lower Nubia. Fields of change: progress in African archaeobotany. Barkhuis & Groningen University Library, Groningen, The Netherlands, 157-164 (2007).

(50) Clapham, A., & Rowley-Conwy, P. The archaeobotany of cotton (Gossypium sp. L.) in Egypt and Nubia with special reference to Qasr Ibrim, Egyptian Nubia. From foragers to farmers. Papers in honour of Gordon C. Hillman. Oxbow Books, Oxford, 244-253 (2009).

(51) Adams, WY. *Ceramic Industries of Medieval Nubia.* University of Kentucky Press (1986)

(52) Hanghøj K, Seguin-Orlando A, Schubert M, Madsen T, Pedersen JS, Willerslev E, Orlando L. Fast, Accurate and Automatic Ancient Nucleosome and Methylation Maps with epiPALEOMIX. *Mol. Biol. Evol.* **33**, 3248-3298 (2016).

(53) Paterson AH *et al.* The Sorghum bicolor genome and the diversification of grasses. *Nature* **457**, 551-556 (2009).

(54) Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nature Methods*. **9**(4), 357-359 (2012).

(55) Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G., Abecasis G., Durbin R. and 1000 Genome Project Data Processing Subgroup The Sequence alignment/map (SAM) format and SAMtools. *Bioinformatics* **25**, 2078-2079 (2009).

(56) Bouyer et al. DNA methylation dynamics during early plant life. *Genome Biology* 18:179 (2017)

(57) Allaby, R. G., Stevens, S., Lucas, L., Maeda, O., & Fuller, D. Q. Geographic mosaics and changing rates of cereal domestication. *Philosophical Transactions of the Royal Society B*, *372*, 20160429. (2017).Clark, R.M., Tavare, S., & Doebley, J. Estimating a Nucleotide Substitution Rate for Maize from Polymorphism at a Major Domestication Locus. MBE 22 (11): 2304-2312 (2005).

(58) Van der Auwera GA, Carneiro M, Hartl C, Poplin R, del Angel G, Levy-Moonshine A, Jordan T, Shakir K, Roazen D, Thibault J, Banks E, Garimella K, Altshuler D, Gabriel S, DePristo M. From FastQ Data to High-Confidence Variant Calls: The Genome Analysis Toolkit Best Practices Pipeline. *Current Protocols In Bioinformatics* **43**,11.10.1-11.10.33 (2013).

(59) da Fonseca RR, Smith BD, Wales N, Cappellini E, Skoglund P, Fumagalli M, Samaniego JA, Carøe C, Ávila-Arcos MC, Hufnagel DE, Korneliussen

TS, Vieira FG, Jakobsson M, Arriaza B, Willerslev E, Nielsen R, Hufford MB, Albrechtsen A, Ross-Ibarra J, Gilbert MT The origin and evolution of maize in the Southwestern United States. *Nature Plants* 1, 14003 (2015).

(60) Koslov AM, Aberer A, Alexandros S. ExaML version 3: a tool for phylogenomic analysis on supercomputers. *Bioinformatics* **31**, 2577-2579 (2015).

(61) Felsenstein J Evolutionary trees from DNA sequences: A maximum likelihood approach. *J Mol Evol* **17**(6), 368-376 (1981).

(62) Stamatakis A RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* **22**(21), 2688-2690 (2006).

(63) Pfeifer B, Wittelsbürger U, Ramos-Onsins S, Lercher M. PopGenome: An Efficient Swiss Army Knife for Population Genomic Analyses in R. *Mol. Biol. Evol.* **31**(7):1929-1936 (2014).

(64) Wolfe KH, Sharpe PM, Li WH. Rates of synonymous substitution in plant nuclear genes. *J. Mol. Evol.* 29:208-211 (1989).

Figure 1. Genomic heterozygosity over time in *S. bicolor* type bicolor (N=7).

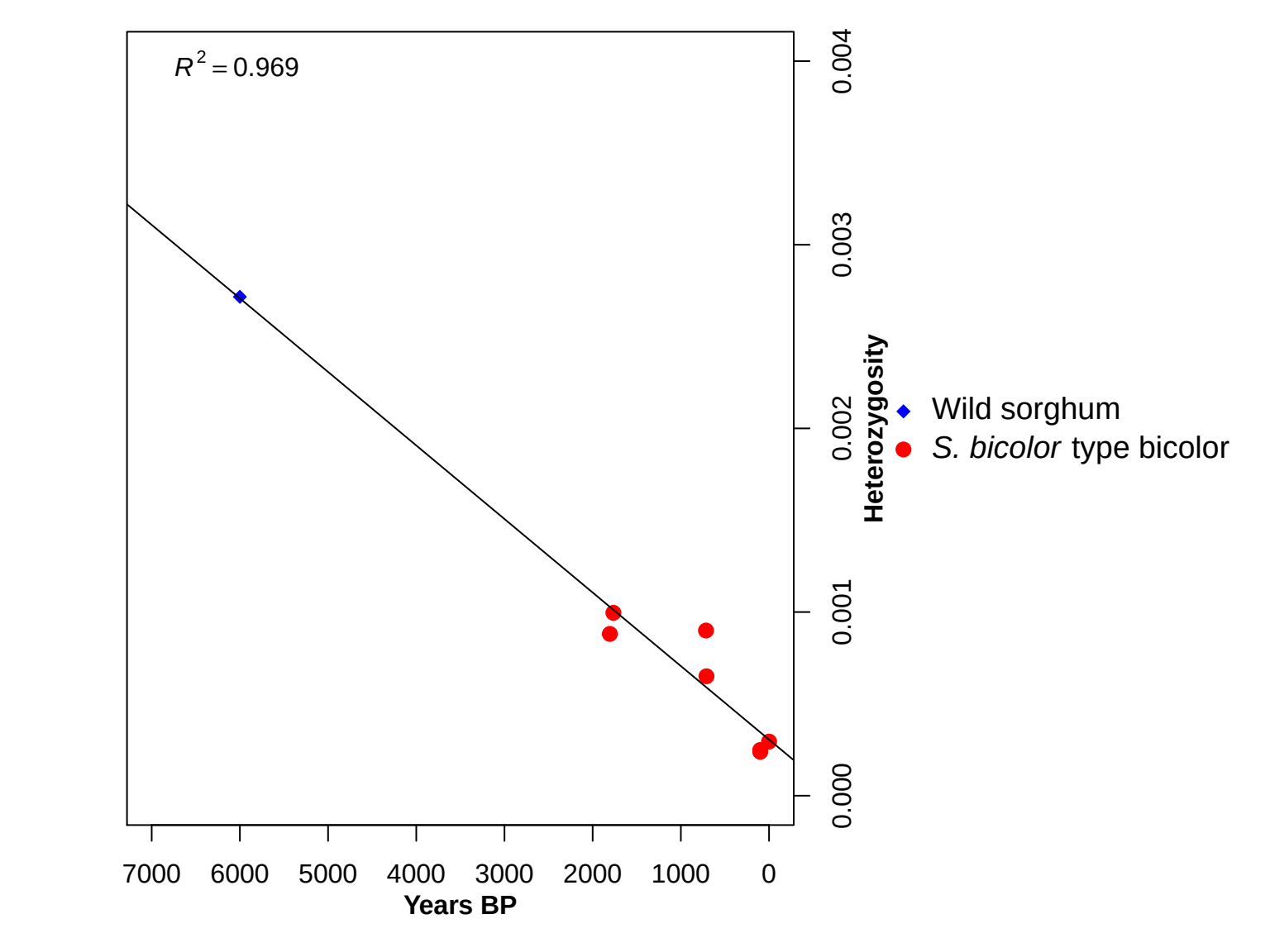
Figure 2. Total recessive GERP load over time in *S. bicolor* type bicolor (N=7).

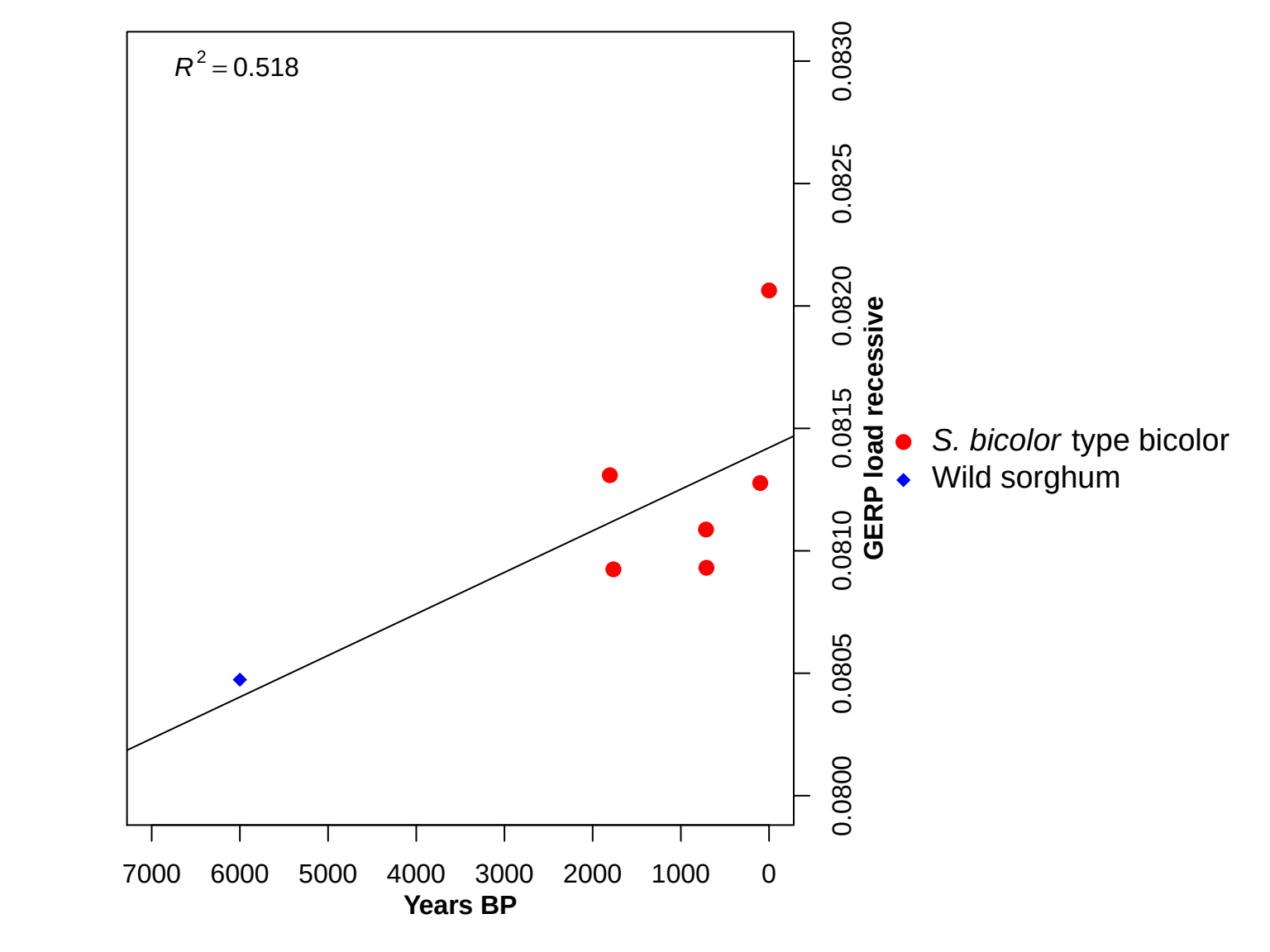
Figure 3. Selection signals across *S. bicolor* chromosomes 1 to 10. Heterozygosity ratio (wild/cultivated) inverted probabilities (Bonferroni corrected) shown in colours as described in key. Grey dashed line indicates 1% significance threshold after Bonferroni correction. SweeD values shown in red. Above: Locations of 38 known domestication genes shown in black. Locations of candidate domestication loci identified by Mace *et al* (24) shown in brown. Locations of GERP score regions of difference (grod) between genomes shown in green.

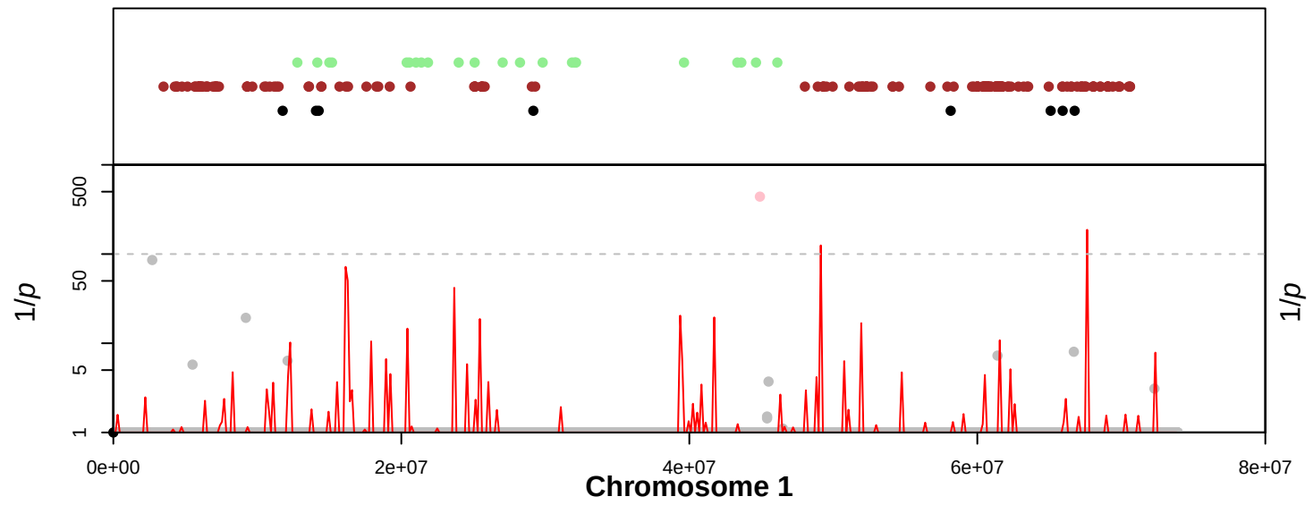
Figure 4. Principal Coordinate Analysis of 1894 SNPs from 23 genomes in this study and 1046 sorghum lines described in Thurber et al (25).

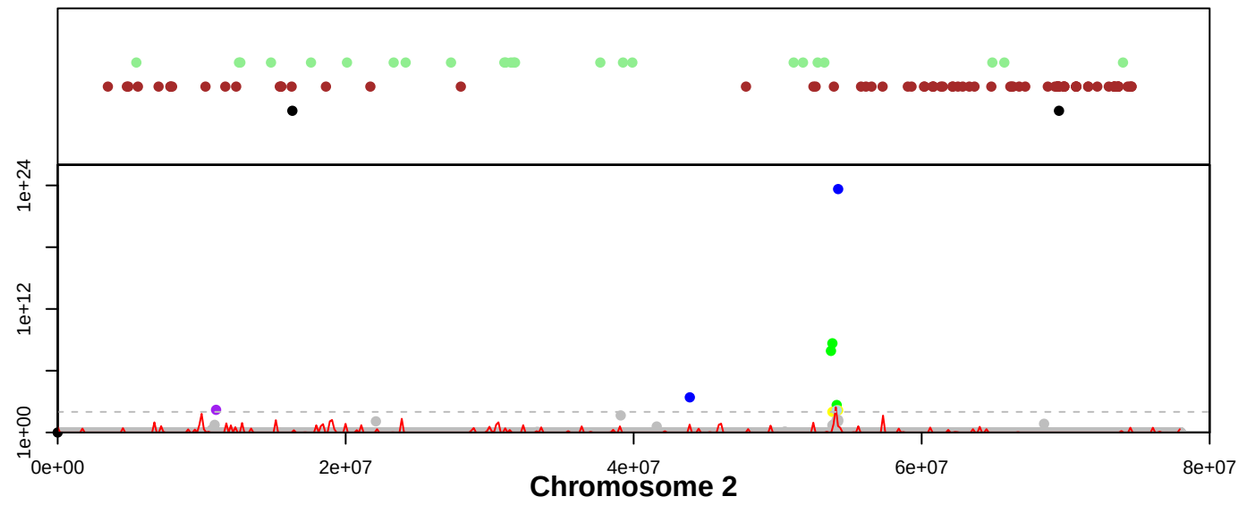
Figure 5. Genome rescue between bicolor and durra lineages. A. Potential genome rescue of descendents from ancestors by donors based on GERP scores. Red indicates the resultant change from combined score of ancestors and donors in regions of observed GERP load reduction in descendents. Blue indicates the genome wide change in gerpGERP score from combining ancestor and donor scores. B. *P* values for the extent of overlap in genomic regions between descendents and potential donors that are lower (*p1*) or higher (*p2*) in

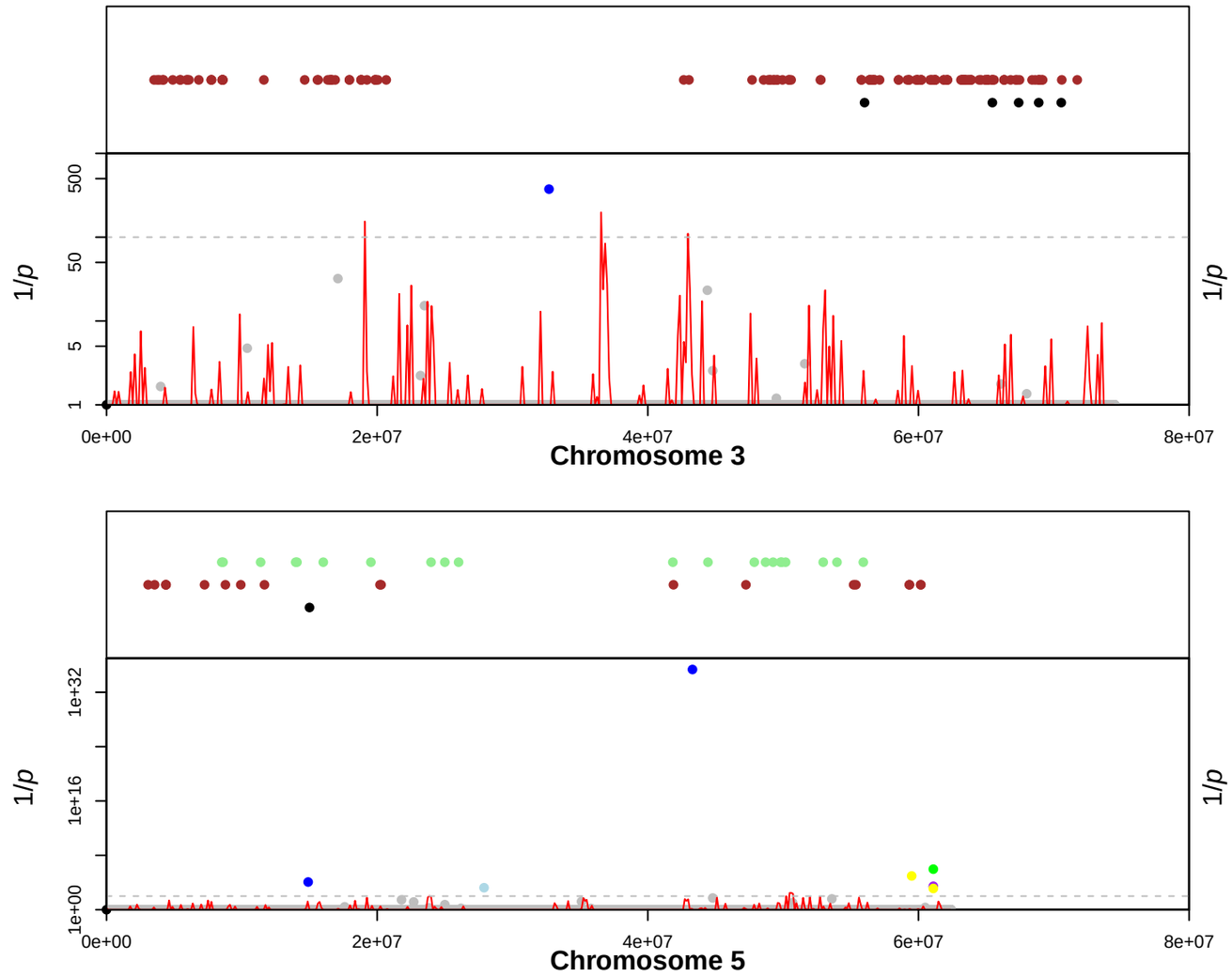
GERP score than the ancestor. *P* values calculated from Beta Distribution overlap of observed and expected proportions of coincidence in mutation load states relative to ancestor (1-tailed test), see methods for details.

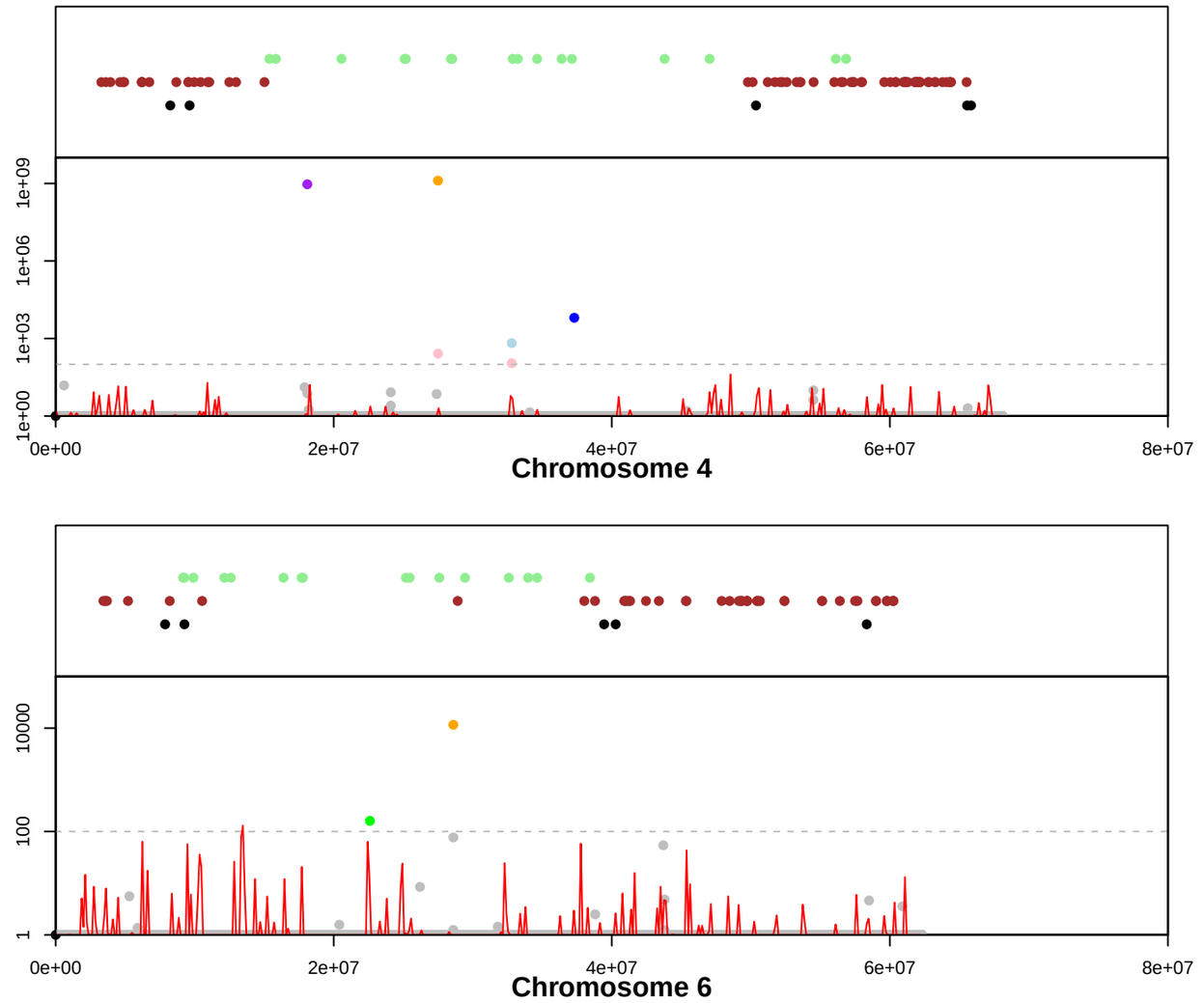

$R^2 = 0.969$
0.004
0.003
0.002
0.001
0.000
Heterozygosity
7000
6000
5000
4000
3000
2000
1000
0
Years BP
Wild sorghum
S. bicolor type bicolor


$R^2 = 0.518$
0.0830
0.0825
0.0820
0.0815
0.0810
0.0805
0.0800
GERP load recessive
7000
6000
5000
4000
3000
2000
1000
0
Years BP
S. bicolor type bicolor
Wild sorghum


1/p
1
5
50
500
0e+00
2e+07
4e+07
6e+07
8e+07
Chromosome 1

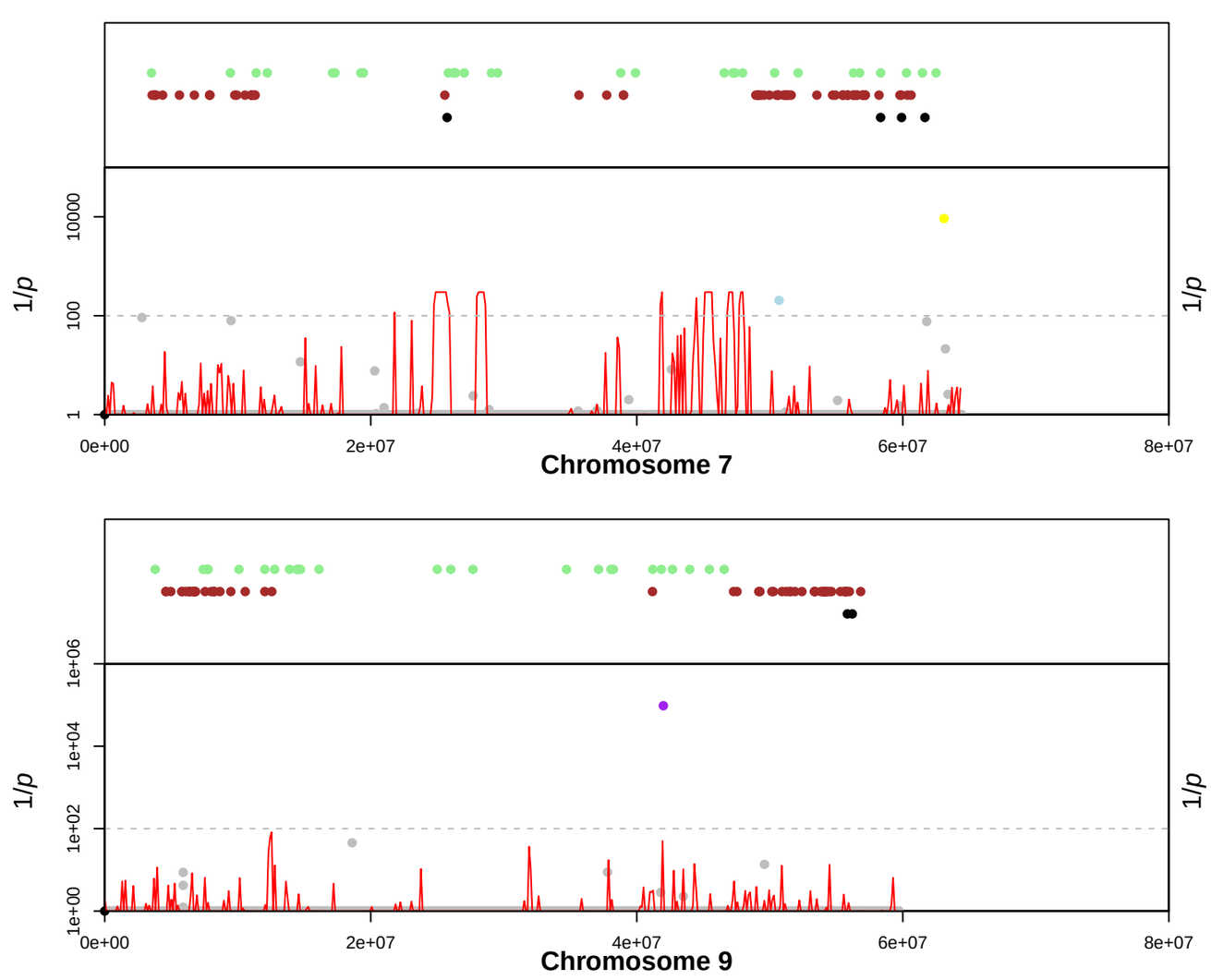

1/p
1e+00
1e+12
1e+24
0e+00
2e+07
4e+07
6e+07
8e+07
Chromosome 2

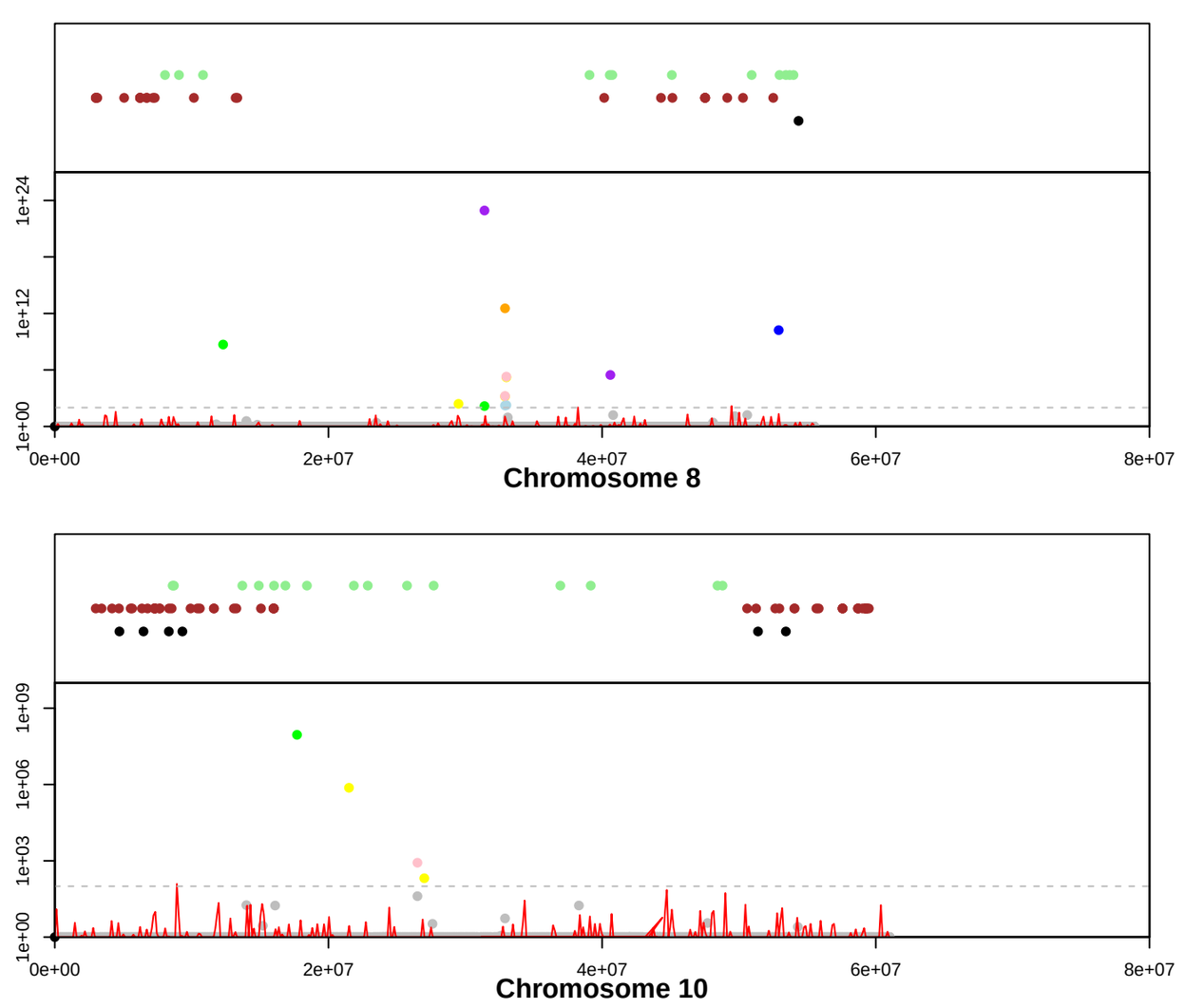

1/p
1
5
50
500
0e+00
2e+07
4e+07
6e+07
8e+07
Chromosome 3


1/p
1e+00
1e+03
1e+06
1e+09
0e+00
2e+07
4e+07
6e+07
8e+07
Chromosome 4

1/p
1e+00
1e+16
1e+32
0e+00
2e+07
4e+07
6e+07
8e+07
Chromosome 5

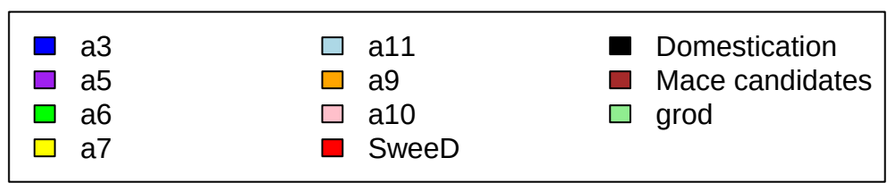
1/p
1
100
10000
0e+00
2e+07
4e+07
6e+07
8e+07
Chromosome 6


1/p
1e+00
1e+02
1e+04
1e+06
0e+00
2e+07
4e+07
6e+07
8e+07
Chromosome 7


1/p
1e+00
1e+12
1e+24
0e+00
2e+07
4e+07
6e+07
8e+07
Chromosome 8

1/p
1e+00
1e+02
1e+04
1e+06
0e+00
2e+07
4e+07
6e+07
8e+07
Chromosome 9

1/p
1e+00
1e+03
1e+06
1e+09
0e+00
2e+07
4e+07
6e+07
8e+07
Chromosome 10


a3
a5
a6
a7
a11
a9
a10
SweeD
Domestication
Mace candidates
grod

PCA via singular value decomposition of 1046 individuals using 1723 markers.

PC2 (VE:4.76%)

PC1 (VE:7.9%)

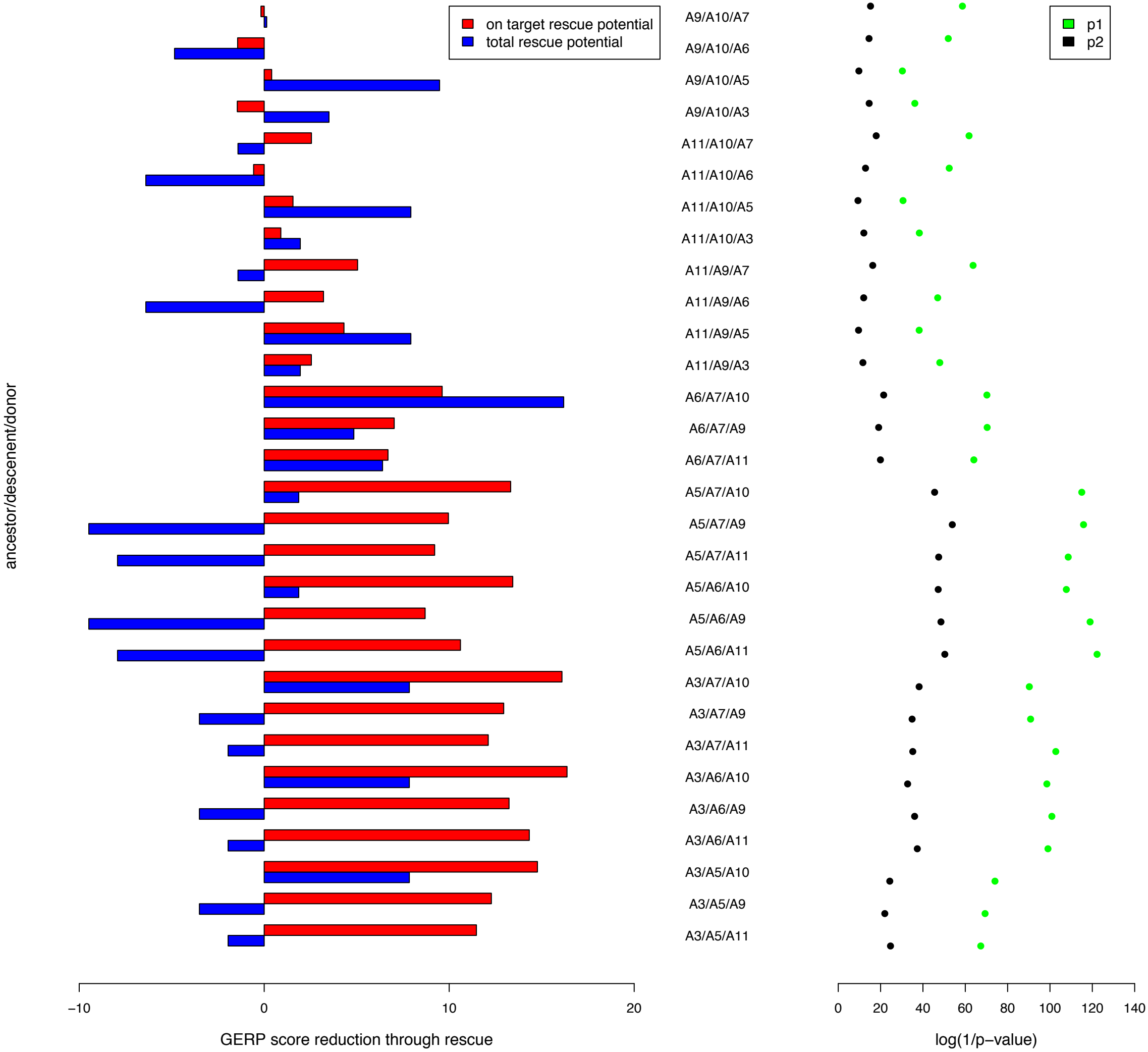
10
0
−10
−20
−10
0
10

M4
M9
M7
M8
M5
M6
A3
M10
A5
M12
Kew2
A8
A6
A7
A9
A10
A11
A12
M11
Kew1
M1
M3
M2

Ancient genomes
Asian durras
Caudatum
E African durras
Guinea
Kafir
Unclassified racial group

(A)

(B)


on target rescue potential
total rescue potential
ancestor/descenent/donor
A9/A10/A7
A9/A10/A6
A9/A10/A5
A9/A10/A3
A11/A10/A7
A11/A10/A6
A11/A10/A5
A11/A10/A3
A11/A9/A7
A11/A9/A6
A11/A9/A5
A11/A9/A3
A6/A7/A10
A6/A7/A9
A6/A7/A11
A5/A7/A10
A5/A7/A9
A5/A7/A11
A5/A6/A10
A5/A6/A9
A5/A6/A11
A3/A7/A10
A3/A7/A9
A3/A7/A11
A3/A6/A10
A3/A6/A9
A3/A6/A11
A3/A5/A10
A3/A5/A9
A3/A5/A11
−10
0
10
20
GERP score reduction through rescue
p1
p2
0
20
40
60
80
100
120
140
log(1/p−value)