The relationship between sexual dimorphism and androgen response element proliferation in primate genomes

Andrew P. Anderson[1,2] (ID) **and Adam G. Jones[3]** (ID)

[1]*Biology Department, Reed College, Portland, Oregon 97202*

[2]*E-mail: aquaticdrew@gmail.com*

[3]*Department of Biology, University of Idaho, Moscow, Idaho 83844*

In the males of many vertebrate species, sexual selection has led to the evolution of sexually dimorphic traits, which often are developmentally controlled by androgen signaling involving androgen response elements (AREs). Evolutionary changes in the number and genomic locations of AREs can modify patterns of receptor regulation and potentially alter gene expression. Here, we use recently sequenced primate genomes to evaluate the hypothesis that the strength of sexual selection is related to the genome-wide number of AREs in a diversifying lineage. In humans, we find a higher incidence of AREs near male-biased genes and androgen-responsive genes when compared to randomly selected genes from the genome. In a set of primates, we find that gains or losses of AREs proximal to genes are correlated with changes in male expression levels and the degree of sex-biased expression of those genes. In a larger set of primates, we find that increases in indicators of sexual selection are correlated with genome-wide ARE counts. Our results suggest that the responsiveness of the genome to androgens in humans and their close relatives has been shaped by sexual selection that arises from competition among males for mating access to females.

KEY WORDS: Androgen, evolution, genome, hormone response elements, primates, sexual selection.

Although it is well established that sexual selection can shape sex-biased phenotypes, there is growing interest in the impacts of sexual selection on the genomic architecture (Mank, 2017, Wilkinson et al., 2015, Kirkpatrick, 2017). As genomic information becomes more readily available, the genome-wide effects of sexual selection may become detectable without *a priori* knowledge of the strength of sexual selection or phenotypic characteristics. Because sexual selection often drives the evolution of sexually dimorphic traits, the extent of sexual dimorphism across species can be used as an indicator of the relative strength of sexual selection, especially in systems where the mating competition is well understood (Dines et al., 2014, Fitzpatrick et al., 2012, Morris and Carrier, 2016). In primates, for example, body mass and canine tooth length are sexually dimorphic in strongly sexually selected species (Plavcan and van Schaik, 1997, Thorén et al., 2006). Additionally, relative testis mass is known to correlate with increased multimale mating behavior and is an indicator of postcopulatory sexual selection, as it is associated with greater sperm counts for sperm competition (Anderson and Dixson, 2002, Fitzpatrick et al., 2012, Kahrl et al., 2016).

Sexual dimorphism is often accompanied by sex-biased gene expression resulting in correlations between expression-level differences and dimorphic features (Harrison et al., 2015, Pointer et al., 2013, Ellegren and Parsch, 2007). Sex-biased gene expression is considered to be one possible resolution to intralocus sexual conflict (Bonduriansky and Chenoweth, 2009, Stewart et al., 2010, Pennell and Morrow, 2013), the situation in which an allele at a locus has different fitness effects depending on the sex of its bearer (Bonduriansky and Chenoweth, 2009, Connallon and Clark, 2014, Lande, 1980). There is evidence that sex-biased regulation via hormones or other signaling pathways can lead to sex-biased gene expression (Mank, 2017, Oliva et al., 2020, Anderson et al., 2020). Recent studies have suggested that sex-biased hormones have pleiotropic effects that reduce genetic correlations

between the sexes (Wittman et al., 2021), and many genes have differential effects on the phenotype depending on the bearing sex (van der Bijl and Mank, 2021). Hence, genomic regions associated with sex-biased signaling could contribute to the phenotypic expression of traits under sexual selection and thus respond to sexual selection.

The most obvious genomic targets of sex-biased hormone regulation are *cis*-regulatory elements, sequences of the DNA usually found in noncoding regions that regulate gene expression (Geserick et al., 2005, Tsai and O'Malley, 1994, Wittkopp and Kalay, 2012). Despite being noncoding regions, they are subject to selection because nucleotide substitutions can affect the strength with which a receptor binds to the genomic region (a value known as binding *affinity*) and how exclusively the genomic region is bound by a particular receptor over others (i.e., the *specificity* of the sequence) (Hahn, 2007, Smith et al., 2013, Wray, 2007). Furthermore, duplications of *cis*-regulatory elements may be important in creating new opportunities for receptor regulation (Jimenez-Delgado et al., 2009) or in altering the transcriptomic response in an additive fashion (Geserick et al., 2005, Cleutjens et al., 1996, Wilson et al., 2016). There is evidence that sex-biased gene expression is altered as physically proximal *cis*-regulatory elements are gained or lost (Matthews et al., 2021). In the human genome, there is a correlation between the number of estrogen-mediated genes and the number of putative estrogen binding sites across chromosomes (Anderson and Jones, 2019). In the Gulf pipefish, there is an excess of estrogen binding regions near putative sex-biased and estrogen-mediated genes involved in ornament expression (Anderson et al., 2020). It is possible that just as gene expression-level differences coincide with sexual dimorphism, signatures of sex-biased regulation via changes in genome-wide *cis*-regulatory elements will likewise coincide with sexual dimorphism.

To investigate the hypothesis that an increase in sexual selection should correspond with an increase in genome-wide regulatory elements, we chose to look specifically in vertebrates because this group has well-documented links between sex-biased signals, sex-biased traits, and the associated *cis*-regulatory elements. Typically, sexual selection acts more strongly in males and leads to male-biased traits, such as ornamentation, weaponry, courtship, and sperm production, which are often regulated by androgens in vertebrates (Bartos et al., 2012, Bhattacharya et al., 2019, Ghosal and Sorensen, 2016, Lindsay et al., 2016). The *cis*-regulatory element associated with androgen receptors is the androgen response element (ARE), a 15 base-pair sequence composed of two palindromic (i.e., reverse complement) half-sites separated by a three base-pair spacer (e.g., AGAACAnnnTGTTCT—half-site, spacer, half-site) (Wilson et al., 2016, Roche et al., 1992) (Table 1). If (1) sexual selection drives the evolution of sexual dimorphism and enlarged testis

size and (2) these features are, at least in part, controlled by sex-biased expression mediated by androgens, then we can predict that new androgen-regulated genetic regions should accumulate over evolutionary time to create new genomic regions with the potential for sex-specific gene expression. These new genomic regions can either occur to induce novel androgen-regulated expression in a gene or enhance an already existing androgen-responsive gene. In either case, the total number of genome-wide AREs would increase. In the present study, we formally tested this prediction by first investigating if AREs are associated with greater sex-biased and androgen-responsive expression. Then we examined correlations between gain or loss of proximal AREs and expression patterns. Finally, we determined the correlation between the number of genome-wide AREs and indicators of both precopulatory (size dimorphism) and postcopulatory (relative testis size) sexual selection. A correlation between the ARE count and the indicators would suggest that sexual selection drives a genome-level change in responsiveness to sex-hormone signaling as adaptations to mating competition arise over evolutionary time.

The ARE nucleotide sequence (motif) with the highest binding affinity to the androgen receptor is called the canonical ARE (*cARE*) (Wilson et al., 2016, Roche et al., 1992). Substituting base pairs in the motif may still allow for androgen binding but will alter the affinity and specificity of the sequence. Although the *cARE* has high affinity for binding to androgen receptors, it has low specificity and will easily bind to other related receptors, especially the glucocorticoid receptor (Laudet et al., 1992, Nelson et al., 1999, Geserick et al., 2003). Paradoxically, higher specificity is achieved via more point mutations on one of the half-sites, so a loss in overall affinity accompanies higher specificity (Wilson et al., 2016, Nelson et al., 1999, Sahu et al., 2014). There are numerous variants of the ARE (Table 1) that have been empirically confirmed. Some of these variants are proposed to be general, whereas others are proposed to be specific to the androgen receptor. Another aim of our study was to see which validated ARE motifs more strongly correlate with indicators of sexual selection and if they share any patterns in nucleotide sequence.

To investigate the emergent patterns in genome-wide counts of various ARE motifs and sex-biased traits associated with sexual selection, we required a vertebrate taxon containing a multitude of species with well-sequenced genomes as well as data on characteristics associated with pre- and postcopulatory sexual selection. We identified simiiform primates as a viable taxonomic group. These primates have been extensively studied in terms of sexual dimorphism, testis size, and genomic sequencing. For sexually dimorphic traits, and therefore indicators of precopulatory sexual selection, we identified body mass dimorphism (Plavcan and van Schaik, 1997, Mitani et al., 1996) and canine height dimorphism (Thorén et al., 2006, Plavcan and van Schaik, 1992) as

Table 1. ARE motifs used in this study with reference (in parentheses) and DNA sequence.

Motif	1st Half-Site						2nd Half-Site						Mass	Canine	Testis
	-6	-5	-4	-3	-2	-1	1	2	3	4	5	6			
NON-SPECIFIC ANDROGEN RESPONSE MOTIFS															
cARE (Wilson & Filipp, 2016))	A	G	A	A	C	A	T	G	T	T	C	T	0.253 (0.048) λ = 0.948	0.688 (0.368) λ = 0.960	0.268 (0.072) λ = 0.975
psa3 (Cleutjens et al., 1997)	G	G	A	A	C	A	T	G	T	A	T	C	**0.070*** (0.008) λ = 1.000	0.700 (0.399) λ = 1.000	**0.284*** (0.068) λ = 1.000
hklk (Murtha et al., 1993)	G	G	A	A	C	A	A	G	T	G	C	T	0.407 (0.179) λ = 0.544	**0.589*** (0.347) λ = 0.000	**0.624** (0.378) λ = 0.224
hk2 (Mitchell et al., 2000)	G	G	A	A	C	A	T	G	T	A	C	C	0.115 (0.013) λ = 0.564	**0.492** (0.228) λ = 0.139	0.211 (0.035) λ = 0.530
psa1 (Riegman et al., 1991)	A	G	A	A	C	A	A	G	T	G	C	T	-0.001 (0.001) λ = 0.987	0.037 (0.024) λ = 1.000	0.121 (0.000) λ = 0.995
c3 (Claessens et al., 1989)	A	G	A	A	C	A	C	G	T	A	C	T	-0.052 (0.004) λ = 1.000	0.133 (0.000) λ = 1.000	0.080 (0.019) λ = 1.000
p21 (Lu et al., 1999)	G	G	A	A	C	C	C	G	T	G	C	T	-0.135 (0.031) λ = 0.330	-0.023 (0.003) λ = 0.281	0.384 (0.113) λ = 0.266
slp3 (Adler et al., 1992)	A	G	A	A	C	A	T	G	T	T	T	C	0.071 (0.000) λ = 0.885	0.403 (0.098) λ = 0.891	0.204 (0.043) λ = 0.860
aARE (Claessens et al., 2017)	G	G	A	A	C	A	T	G	T	T	C	C	**-0.242** (0.036) λ = 0.951	0.381 (0.118) λ = 1.000	0.324 (0.088) λ = 0.999
cSRE (Roche et al., 1992)	A	G	A	A	C	A	T	G	T	A	C	C	-0.146 (0.022) λ = 0.867	**-0.185** (0.099) λ = 1.000	-0.278 (0.053) λ = 0.830
aSRE (Roche et al., 1992)	A	G	A	A	C	A	T	G	T	T	C	C	0.118 (0.001) λ = 0.862	0.345 (0.105) λ = 0..851	0.199 (0.016) λ = 0.830
mvdp (Fabre et al., 1994)	A	G	A	A	C	A	A	C	T	T	C	A	-0.076 (0.011) λ = 0.000	-0.202 (0.041) λ = 0.000	-0.196 (0.039) λ = 0.000
ANDROGEN-SPECIFIC RESPONSE MOTIFS															
ccnd1 (Lanzino et al. 2010)	A	G	A	A	C	T	T	T	A	G	C	A	-0.007 (0.000) λ = 0.254	**0.417*** (0.174) λ = 0.000	0.255 (0.055) λ = 0.097
sc12 (Verrijdt et al., 1999)	A	G	A	A	C	T	A	G	A	G	C	C	0.290 (0.043) λ = 0.964	0.624 (0.235) λ = 0.902	0.477 (0.184) λ = 0.928
pem2 (Barbulescu et al., 2001)	A	G	C	A	C	A	T	G	C	T	C	A	0.247 (0.033) λ = 0.920	0.639 (0.338) λ = 0.804	0.120 (0.014) λ = 0.917
pem1 (Barbulescu et al., 2001)	G	G	A	A	C	A	A	G	A	T	C	T	-0.191 (0.012) λ = 1.000	0.183 (0.098) λ = 1.000	0.197 (0.019) λ = 1.000
pb2 (Claessens et al., 1996)	A	G	T	A	C	T	A	G	A	A	C	C	-0.014 (0.018) λ = 0.908	0.355 (0.076) λ = 0.903	0.073 (0.000) λ = 0.902
slp2 (Verrijdt et al., 2000)	A	G	A	A	C	T	T	G	A	C	C	A	-0.036 (0.004) λ = 0.982	-0.234 (0.126) λ = 0.980	-0.130 (0.008) λ = 0.987
PC ARE													0.028 (0.001) λ = 0.501	**0.494*** (0.244) λ = 0.000	**0.279*** (0.370) λ = 1.000
							# with positive correlation						8	**14**	**15**

Colors are used to better visualize shared nucleotides, and the 3bp spacers between half-sites are not shown. Numbers under the half-site indicate the base pair position referenced in the text. For each motif, we report the results of the GLS model of motif count with each trait. We report the correlation (R^2 value) and phylogenetic signal (λ). Significant values are in bold and significant values after Bonferroni correction have an asterisk. PC_ARE represents the first component from the principal components analysis for all response elements.

viable characteristics for analysis. In addition, we identified relative testis mass as an indicator of postcopulatory sexual selection (Anderson and Dixson, 2002, Harcourt et al., 1995, Moller, 1988). We also selected 18 variations of the ARE motif, including six androgen-specific ARE motifs and 12 nonspecific ARE motifs. Of the 12 nonspecific motifs, we included the canonical ARE (*cARE*), as well as an alternate to the canonical ARE (*aARE*) that differs at the end base pairs only (Claessens et al., 2017). We also considered two versions of the generic steroid response element: the canonical SRE (*cSRE*) and an alternate SRE (*aSRE*), which bind generally to multiple corticosteroids (Roche et al., 1992). To investigate correlations between AREs and expression

levels, we used sex-biased (Guo et al., 2018) and androgen responsive data from humans (Nelson et al., 2002, Romanuik et al., 2009) as well as expression data from five organs (Brawand et al., 2011).

Here, we explore two dimensions of AREs and sex bias. In the initial line of questioning, we test whether gene proximity to AREs is greater in male-biased and androgen responsive genes relative to all genes. In a small sample of hominid primates, we tested for correlations between changes in male expression levels and sex bias for select genes, and we examine the gain or loss of AREs proximal to those genes. In the second line of questioning, we tested multiple hypotheses regarding the relationship between the number of genome-wide ARE motifs, the different motif sequences, and the three traits related to sexual selection in simiiform primates. First, we test for a significant correlation between genome-wide counts of each ARE motif and each of the physical traits. Second, we employ multidimensional analysis to look for clusters of ARE motifs that group with respect to physical traits. After determining which, if any, ARE motifs show a positive relationship with physical traits, we then investigate possible patterns of sequence or androgen specificity among those motifs.

Methods
DATA COLLECTION

Thirty-one species of simiiform primate had published genomes at the time of this study. To evaluate the completeness of genomic sequencing, we used BUSCO (Simão et al., 2015) and selected primate genomes that contained more than 80% of the single-copy orthologous gene set from the primate database as complete sequences. After this selection step, we were left with 26 primate species for our study. Using existing literature, we gathered data on body mass, canine height, and testis weight for these species (Thoren et al., 2006, Harcourt et al., 1995, 2017, Plavcan and Ruff, 2008) (Table S1). In some cases, data were available from a congener or were gathered prior to a taxonomic revision, splitting a formerly single species into multiple species. For *Nomascus leucogenys*, testis dimensions were taken (Hill and Kanagasuntheram, 1959) and used to estimate testis mass (Dahl et al., 1993). For three species (*Rhinopithecus roxellana*, *Cercopithecus mona*, and *Piliocolobus tephrosceles*), no comparable data were available for testis weight. Both body mass and canine height dimorphism were calculated using the residuals from a fitted log-log model of male and female values for each species (Abouheif and Fairbairn, 1997, Clutton-Brock, 1985, Rensch, 1950) such that larger residuals represent greater male-biased dimorphism. Similarly, relative testis mass was calculated using the residuals from a fitted log-log model of body and testis mass for each species (Moller, 1988) such that larger residuals represent greater relative testis mass.

To investigate genes and their association with AREs, sex bias, and androgen responsiveness, we used three datasets: one that reported a list of sex-biased genes in humans (Guo et al., 2018) (Table S2), and two that reported genes and expression-level changes after exposure to androgen in human LNCaP cell lines (Nelson et al., 2002, Romanuik et al., 2009) (Tables S3 and S4). For expression data, we used data found in five species (*Homo sapiens*, *Pan troglodytes*, *Pan paniscus*, *Gorilla gorilla*, and *Macaca mulatta*) across five organs (brain, cerebellum, heart, kidney, and testis) and numerous genes (Brawand et al., 2011). In many cases, the datasets contained just one replicate, but where multiple individuals were sampled within a species, we took the arithmetic mean of the RPKM (Reads per kilobase of transcript per million mapped reads). We used RPKM values of males as expression values and calculated sex bias by determining the $\log_2$-fold difference between male and female values within a species. Given the small sample sizes (mostly $n = 1$), we did not calculate significance of sex bias.

We selected ARE motifs for analysis by searching the literature for research that established the role of the element in a transcriptional response to androgen for a given gene (Table 1). This could have been done by excision of the motif in question or transfection of the motif to generate a response to an androgen. We limited our dataset to 18 motifs for ease of multidimensional analysis. All motifs were described using either mouse or human cell lines making them appropriate for study in our primate group. For each species, we searched for each motif in the genome using the *exactmatch* feature in the R library Biostrings (Pagès et al., 2019). This feature counts the number of times a set of nucleotides has the exact pattern of base pairs in the scanned sequence, in our case the genome, as the given motif. Because our chosen motifs often diverge by one to two base pairs, we did not allow any mismatches in the motif search. Scans were done only in one direction, which was sufficient to identify all sites for palindromic motifs. For nonpalindromic motifs, we ran a second scan with the reverse complement of the motif and summed the number of matches to represent total hits in the genome. We used the *phytools* (Revell, 2012) heatmap function to calculate a standardized score (Z-score) of both the genome-wide count data for each motif and residual values for all three indicator traits (Fig. 1; raw data, Table S5). It is important to note that direct detection of a motif should not be equated with the identification of an active androgen receptor binding site (Wasserman and Sandelin, 2004), as many factors contribute to androgen receptor binding. Rather, our goal was to enumerate the total number of motifs in the genome, and these sites should be interpreted as putative regions for androgen receptor binding. In cases where we needed to match ARE location with a gene of interest, we used the General Feature Format (GFF) files associated with the genomes for the five species described in the expression data section.
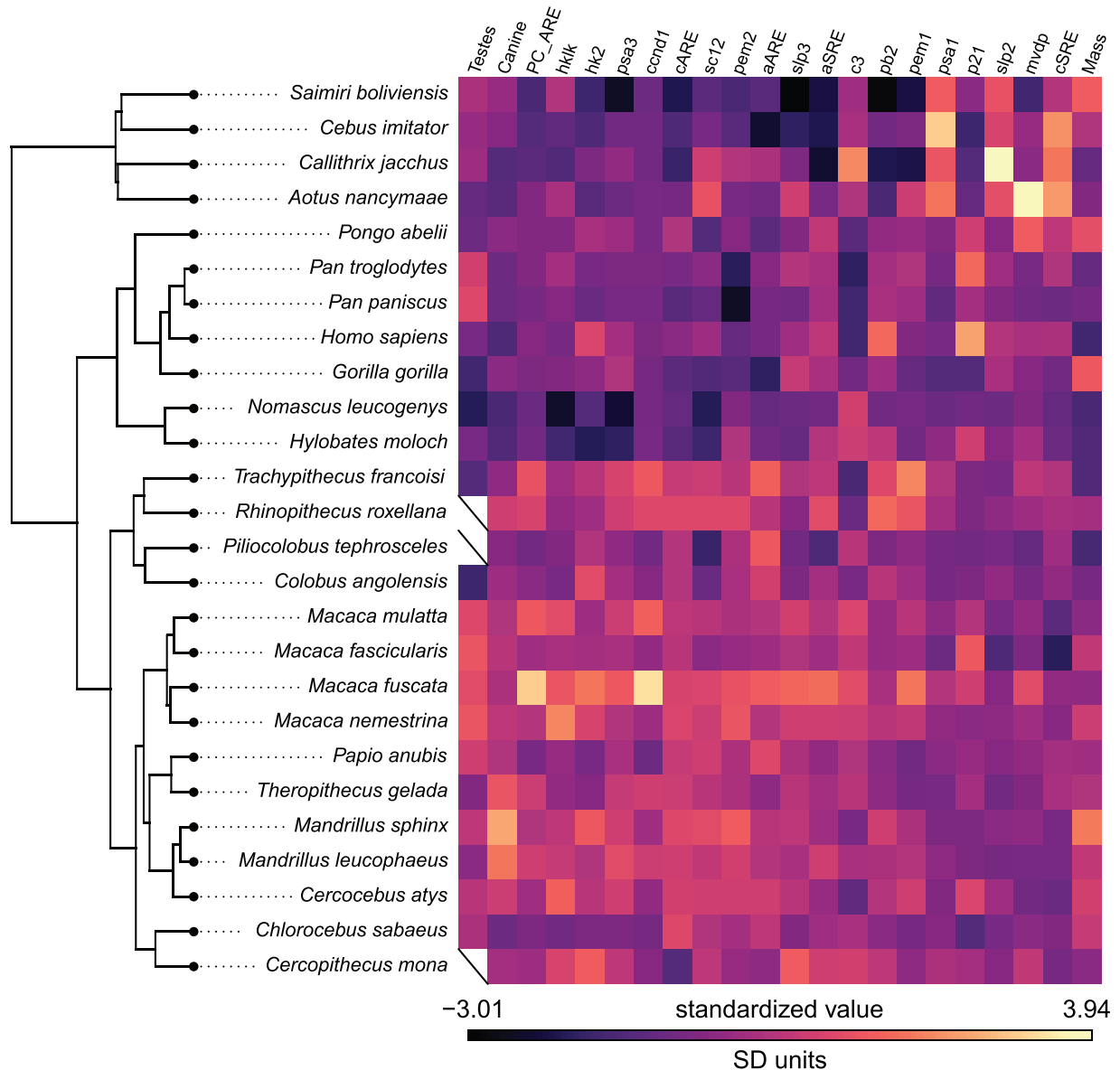
Figure 1. Heat map of standardized values (*Z*-scores) of genome-wide counts of motifs, measures of sexual dimorphism in body mass and canine size, and measure of relative testis size by study species in the simiiform primates, with accompanying phylogeny. These values were calculated by using the heatmap function in phylools (which standardizes each set of values to a mean of 0 and a standard deviation of 1). Data are missing for testis size in *R. roxellana*, *P. tephrosceles*, and *C. mona*.

ANALYSIS OF SEX BIAS, ANDROGEN RESPONSIVENESS, AND ARES IN HUMANS

We used a custom R script to extract gene coordinates from the human GFF file and then selected genes from this list that matched those that were male-biased from the list of sex-biased genes (*n* = 1043). We then ran the *exactmatch* feature to produce a list of coordinates from the investigated AREs. We used a previously published custom R script (Anderson et al., 2020, Anderson and Jones, 2019) that matches AREs to genes based on a predetermined distance from the transcription start and end sites. In this case, we search 25-kb upstream of the transcription start site, 10-kb downstream of the transcription end site, and the entirety of the gene (both introns and exons). We chose this distance because, although *cis*-regulatory elements can be found far away from the transcription start site (Carroll et al., 2005), the majority are within 25 kb and can also be found in intragenic regions (Zheng et al., 2016, Welboren et al., 2009). To see if the male-biased gene set had more AREs than expected, we randomly sampled an equal number of genes (*n* = 1043) from all genes in the human genome 1000 times to generate two distributions. One distribution is the number of genes with at least one ARE within the predetermined distance (25 kb up, 10 kb down); the other

distribution is the total number of proximal AREs in the entire set. We then determined where the male-biased gene set falls on these distributions. We repeated this process for the androgen-responsive gene list ($n = 122$). If AREs and genes with proximal AREs are overrepresented in the male-biased and androgen-responsive gene sets compared to the random distribution, then there is support for the importance of AREs in male-biased expression and androgen responsiveness, respectively.

CORRELATION OF PROXIMAL ARES AND GENE EXPRESSION

We identified 8250 orthologous genes with complete data across the five primates used for expression analysis. As with humans (described above), we generated a list of proximal AREs and their corresponding genes using the same distances for each species (Table S6). We then filtered genes to keep those with a proximal ARE present in at least one species (leaving 1312 genes). For each organ, we filtered genes so that we kept any gene with an RPKM value in males greater than 10 and an absolute male-female $\log_2$-fold difference greater than 2 in at least one of the five species. In testis, we only used the RPKM value cutoff as we did not have ovary data to compare. We chose these cutoffs a priori to investigate those genes with biological significance and to account for errors due to low sample size. To perform a correlation analysis, we removed any gene without any change in ARE count across species. We downloaded the ultrametric tree for our selected species from the 10k trees website (Arnold et al., 2010) to use as our phylogeny. We created three models using generalized least squares (GLS) with the *phytools* package (Revell, 2012) in R, where both RPKM value in males and sex bias were independently treated as a function of the number of proximal AREs for each gene. The models were (1) no phylogenetic effect on the model, (2) including a phylogenetic effect (*corPagel*) where the phylogenetic effect's value (lambda) was freely estimated in the model, and (3) using a predetermined value of lambda, estimated using the *phylosig* function. We compared the Akaike Information Criterion (AIC) (Akaike, 1974) of the three models and chose the model with the lowest AIC value. The *P*-values from the set of RPKM or sex-bias analyses for each organ were adjusted for multiple tests using the False Discovery Rate (FDR). We further estimated the correlation values by estimating the covariance with lambda set to the value determined by the appropriate model (no phylogenetic effect means lambda = 0) and converted those covariances to correlations. We calculated R^2 from the chosen model.

To test for overrepresentation of any ARE motifs, we calculated a genome-wide baseline for each motif by counting the number of hits in the genome in each species and taking the average of all five species to represent the genome-wide group. For the proximal ARE group, we counted the number of each mo-

tif present in the set of 1312 genes. Because AREs were mostly conserved across species within genes, we only counted a motif once per gene. With a count table for genome-wide and proximal AREs for each motif, we performed a chi-square analysis to test for significantly different distributions. We then investigated the residuals for each motif to determine the percent contribution of each motif to the chi-square statistic. If any motif had a large contribution to the differences between the two sets, we removed it and ran the chi-square test again to confirm its effect.

INDIVIDUAL CORRELATION ANALYSIS

We downloaded the ultrametric tree for all of our 26 selected species from the 10k trees website to use as our phylogeny. As before, we created the three models using generalized least squares where the genome-wide count of an ARE motif was treated as a function of each indicator of sexual selection. Using the chosen model, we looked at the significance of the effect of indicator on motif number both with and without Bonferroni correction for multiple tests of each indicator. As in the previous analysis, we estimated the correlation values and calculated R^2 from the chosen model.

MULTIDIMENSIONAL ANALYSIS

To assess larger patterns of trait correlations across all ARE motif counts, we used several statistical approaches. First, we investigated whether ARE motif counts were more likely to be positively correlated with indicators of sexual selection by using a binomial test for the number of positively correlated motif counts with a given indicator, where the expected number by chance was one half of the total number of motif counts ($n_{Exp} = 9$). Second, we performed a principal components analysis and identified the principal component that had the largest number of loadings with the same directionality and explained a large amount of variance. We then took that principal component (*PC ARE*) and did the above-described GLS modeling to determine its correlation with indicators of sexual selection. Using the 23 species with data available for all three indicator traits, we used canonical correlation analysis (CCorA) (Harold, 1936) to maximize correlations between ARE motif counts and indicator traits. CCorA is symmetrical in that indicator traits and ARE motif count correlations are maximized in both directions. We were able to account for phylogeny in the CCorA using canonical correlation in the *phytools* package. We used both *NbClust* (Malika et al., 2014) and *mclust* (Scrucca et al., 2016) to determine optimal numbers of clusters. We then used kmeans clustering for the number of optimal clusters to determine groupings of ARE motifs in the CCorA.

MOTIF SEQUENCE PATTERN

We further attempted to determine if there was some uniformity to ARE motif sequences that either significantly positively
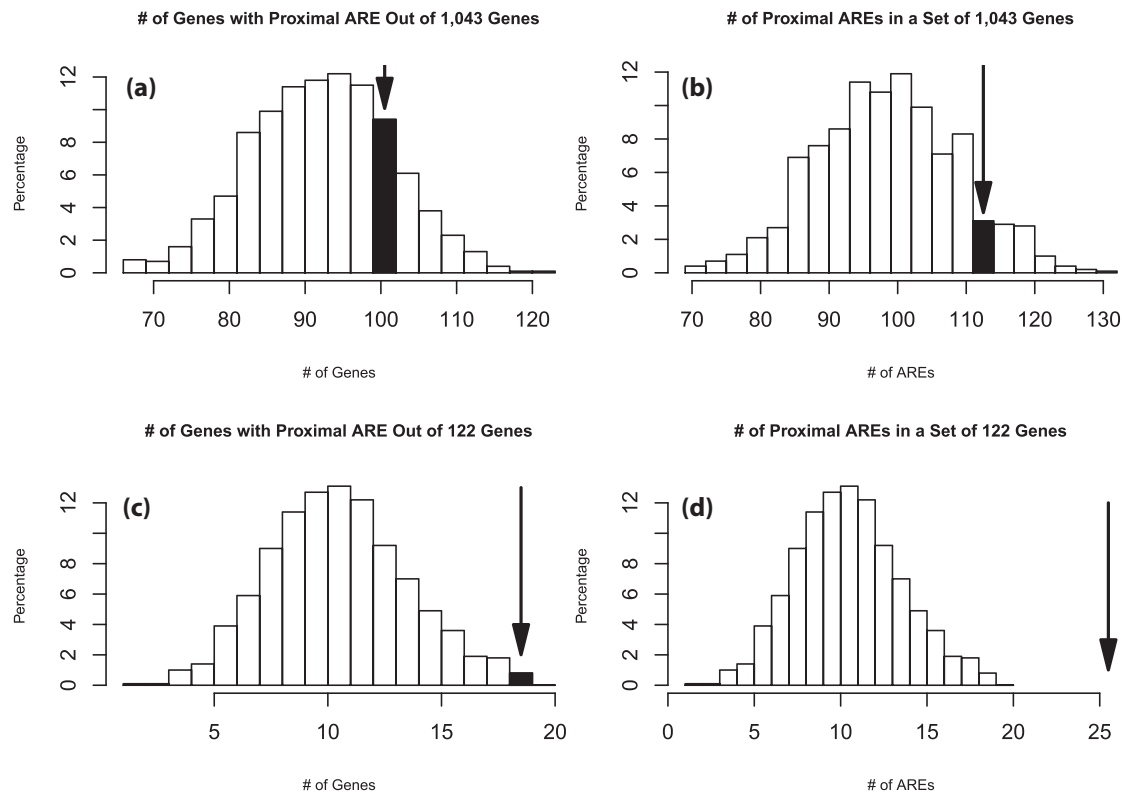
Figure 2. Histograms using distributions from randomly drawn genes (a and b, *n* = 1043; c and d, *n* = 122) from the entire *Homo sapiens* genome taken 1000 times. For each random set of genes, the number of genes with at least one proximal ARE (a and c) and the number of total proximal AREs in the set were used to generate the distributions. The values for the set of male-biased genes (a and b) and androgen-responsive genes (c and d) are indicated with dark bars and arrows, demonstrating that this set is at the upper end of the distributions.

correlated with indicator traits in the GLS analysis or showed positive loadings in the multidimensional analyses. These ARE motif counts are putatively informative about the strength of sexual selection. We generated a haplotype network for the motifs using *pegas* (Paradis, 2010) to visualize similarities and differences among motifs. We also compared the frequency of each base pair of the putatively informative motifs to those motifs that showed no correlations in our analyses described above. It has been suggested that the loss of thymine at position 3 in the motif confers higher specificity to the androgen receptor (Nelson et al., 1999, Claessens et al., 2017, Verrijdt et al., 2003). Seventeen of the 18 motifs chosen have been tested for androgen specificity with competitive assays (Nelson et al., 1999, Geserick et al., 2003, Denayer et al., 2010, Schauwaers et al., 2007) except for *ccnd1*, which has an adenine at the third position. Six of these motifs are androgen specific and lack thymine at the third position. We further investigated whether these androgen-specific motifs are more likely to correlate with indicator traits.

Results

EFFECT OF ARES ON SEX BIAS AND EXPRESSION

Of the 1043 genes with male sex bias in humans, 101 have at least one proximal ARE. In the entire gene set, 111 proximal AREs were detected. These values are in the 80.0 and 87.1 percentiles, respectively, of 1000 random draws of an equal number of genes across the human genome (Fig. 2a,b). These numbers suggest some possible influence of AREs on sex bias but not a significant one. For the 122 androgen-responsive genes in human LNaCP cell lines, 18 have at least one proximal ARE, and in the entire gene set, a total of 25 proximal AREs are present. These values are in the 97.5 and 100.0 percentile, respectively, of 1000 random draws of an equal number of genes across the human genome (Fig. 2c,d). These numbers provide clear evidence that androgen-responsive genes have an excess of AREs.

For the comparative analysis across the five primate species, of the 1312 genes have at least one ARE in at least one species, and 946 of these genes passed our organ-based RPKM filter. Of these 946 genes, 431 show a significant correlation between the

Table 2. Genes with correlations between the number of proximal AREs and gene transcription measures for five primate species across five organs, as described in the text.

Organ	Measure	# Genes Tested	# Genes FDR<0.05	% Genes FDR<0.05	List of Significant Genes
Brain	Male RPKM	64	15	23.44%	CLVS2, *TNNT2*, **SUPT3H**, BST1, *CHADL*, *ESRRG*, NTNG2, **TMEM273**, **RESP18**, PLEKHM3, ***LAPTM5***, SCAND1, *TNFRSF17*, ***RPS29***, **FAM151B**
	Sex-bias	64	9	14.06%	*SLC27A3*, ***SUPT3H***, ***RPS29***, **TMEM273**, EGR2, **FAM151B**, ***LAPTM5***, **RESP18**, WFDC5
Cerebellum	Male RPKM	58	14	24.14%	ZNF19, **ZNF444**, ANXA1, *CLSTN2*, **SRD5A3**, **KLK5**, **NPFFR1**, *LAPTM5*, KCNG1, **IQGAP2**, *TNFRSF17*, NKD2, AFAP1L1, ***RPS29***
	Sex-bias	58	16	27.59%	**ZNF444**, RERG, DYRK4, **SRD5A3**, **IQGAP2**, PDZRN4, *RBP4*, *BCAS1*, ANO6, **KLK5**, *ALDH3A1*, *CHADL*, BCL11A, CNDP1, **NPFFR1**, ***RPS29***
Heart	Male RPKM	151	44	29.14%	**IL16**, **MMRN1**, ESR2, CGN, **LMOD1**, ADGRG3, ***TAGLN***, **EPHX1**, **IL33**, DNM1, ST14, *CHADL*, CPNE4, ***IFIT3***, *CLSTN2*, *TNFRSF17*, DPH5, SLC9A3, ACTN1, **PDE1C**, **ADAM23**, NR4A3, **FASN**, ***ESRRG***, ***PARP9***, *SLC27A3*, *CPVL*, TRIM29, **BBX**, CCDC77, TBC1D2, ARMC1, CD274, MAPK7, ***ASB15***, **SV2A**, TATDN3, PCDH7, GPM6A, *FCER1G*. **SLC19A2**. UNC5A. *SUPT3H*. RBBP9
	Sex-bias	151	43	28.48%	PDGFRL, **MMRN1**, MYL3, MAPK10, **EPHX1**, AAK1, GALC, EPHB6, *TNNT2*, ***TAGLN***, **LMOD1**, ***ESRRG***, FRMD6, *NEURL1B*, **BBX**, **IL33**, HECW2, DGKG, PLCL1, **ADAM23**, GCC2, **PDE1C**, LANCL2, **SLC19A2**, **FASN**, **IL16**, DPH6, *ALDH3A1*, XDH, ATP5PF, PRRG3, ***ASB15***, CDK1, CARTPT, CSF3, ***IFIT3***, CCBE1, ***PARP9***, *RBP4*, ZDHHC2, **SV2A**. SLC7A10. LNX1
Kidney	Male RPKM	68	23	33.82%	SLCO1A2, MCM10, LAMB4, PARP15, RIPPLY1, *BCAS1*, ASPG, **DNTT**, *FCER1G*, **SLC2A9**, ***NEURL1B***, ERP27, **GPX2**, *IFIT3*, *PARP9*, *CPVL*, *ASB15*, ***TAGLN3***, *NR4A3*, SLC22A16, *TNFRSF17*, ATF7IP2, ABCA1
	Sex-bias	68	9	13.24%	TEX26, CCDC42, CRTAC1, ***TAGLN3***, **SLC2A9**, ***NEURL1B***, PER1, SLC37A1, **GPX2**, **DNTT**
Testis	Male RPKM	847	348	41.09%	See Supplemental Table 7

We filtered genes from the initial list of 8250 genes by the presence of a proximal ARE, a male RPKM value >10, and a male to female $\log_2$ foldchange >2 (sex bias), for at least one species in each category. The number of genes passing filter is shown, along with the number of genes with significant correlations. Human gene symbols for significant genes are shown. Bold genes are significant in both male RPKM and sex bias for the same organ. Italicized genes are significant in two separate organs. The list of testis genes can be found in Table S7.

number of proximal AREs and male expression levels or degree of sex-bias expression (or both) for at least one organ (Tables 2 and S7). Most of the significant correlations occur in testis because there was one less level of filtering (no sex-bias criterion); however, 41% of all genes tested in testis show a significant correlation between male expression values and the number of proximal AREs. The remaining organs tend to have more significant correlations with male expression values (23–33% of genes tested) than with sex bias (13–28%). Roughly 40% of genes with sex-bias correlations to proximal ARE numbers also have male expression correlations to proximal ARE numbers. R^2 values are typically high (i.e., >0.9) for significantly correlated genes, whereas correlation coefficients range from 0.1 to 0.99

(Table S7). These data support the supposition that altering the number of local AREs can affect male and sex-biased expression for a gene.

Tests for over- and underrepresentation of specific AREs reveal the motif *mvdp* is strongly underrepresented for both the gene set with at least one proximal ARE and the gene set that passed the filters. In both cases, the residuals of *mvdp* strongly influence the distribution (~20%) and removal of *mvdp* from the lists results in a lack of significance in the chi-square tests (with *mvdp*: $P = 0.0046$, without: $P = 0.0607$). For the gene set with significant correlations between male expression values or sex bias and the number of proximal AREs, the motif *p21* was strongly overrepresented, as the residuals greatly influence the

distribution (47%) and removal of *p21* from the list results in a lack of significance in the chi-square test (with *p21*: $P = 0.0249$, without: $P = 0.5128$). Interestingly, *mvdp* has the most genome-wide hits of any tested motif (range: 988–1255) and *p21* has the fewest (range 47–66).

EFFECT OF SEXUALLY SELECTED TRAITS ON ARE MOTIF NUMBERS

Of the 18 studied ARE motifs, four motifs have at least one significant positive correlation between their genome-wide counts and an indicator trait (Table 1). For these significant positive correlations, canine size dimorphism correlates with three motif counts (*hklk*, *ccnd1*, *hk2*), relative testis mass correlates with two (*hklk*, *psa3*), and body mass dimorphisms correlates with just one (*psa3*). The *hklk* motif count has a good model fit and a high correlation with both testis mass and canine dimorphism, as do *hk2* and *ccnd1* motif counts with canine dimorphism. Looking at the correlations between the genome-wide motif counts and canine size, many motif counts have large correlation values (e.g., *cARE*, *sc12*, *pem2*) but lack statistical significance. The genome-wide count of *psa3* has a poor model fit and low correlation with body mass dimorphism and testis mass, whereas canine size has both high fit and correlation yet is not significant. Looking across indicator traits, motif counts with high correlational values in one trait tend to have high values in the other traits, even if these correlations are not significant in all or even one trait, especially when looking at canine size and testis mass. Two motif counts (*cSRE* and *aARE*) have a significant negative correlation with an indicator trait (canine dimorphism and body mass dimorphism, respectively), but both have poor model fit and correlation values close to zero. Three motif counts have negative correlations with all three indicator traits: *cSRE*, *mvdp*, and *slp2*. In general, for body mass dimorphism the models describe low correlation values and poor model fit with motif counts, whereas for testis mass and canine dimorphism the models tend to have higher correlation values and better model fit with motif counts.

Using the binomial test, where the null expectation of positively correlated motif counts is half the total sampled motifs ($n_{Exp} = 9$), we find that more motif counts positively correlate with canine dimorphism ($n_{Obs} = 14$, $P = 0.030$) and relative testis mass ($n_{Obs} = 15$, $P = 0.008$) than expected by chance, but no excess of positive correlations exists for body mass ($n_{Obs} = 8$, $P = 0.815$). Principle components analysis of the ARE motif counts finds the first principal component explains 52.7% of the variance and has all but three motif counts (*cSRE*, *psa1*, and *slp2*) with positive loadings; two of these three motif counts have negative correlations with all three traits. No other component has as many loadings with the same directionality. This component (*PC ARE*) positively correlates with both canine size dimorphism and relative testis mass with a good fit and high

correlation but does not correlate with body size dimorphism (Table 1). Because the *PC ARE* represents a general increase in motif count (indicated by the large number of positive loadings), it supports the finding of more positive correlations across motif counts than expected by chance. The correlation value for *PC ARE* and canine dimorphism is higher than that for *PC ARE* and testis mass, an observation that agrees with the general trend of higher correlations across individual motif counts for canine dimorphism. These results support the hypothesis that as genome-wide ARE counts increase generally across motifs, there is a corresponding increase in some indicators of sexual selection.

MULTIVARIATE ANALYSIS

Phylogenic signal in the CCorA (Fig. 3) is estimated to be near zero ($\lambda = 0.000066$) and the first two canonical variates for each matrix (ARE count and indicator trait) have correlation values greater than 0.96. Pillai's trace *F*-statistic indicates that dimensions one through three are significant ($P = 0.0001$), but dimensions two through three ($P = 0.1097$) and three by itself ($P = 0.5930$) are not. These results indicate a significant correlation between the indicator trait dataset and the genome-wide ARE count dataset. Analyzing the CCorA without accounting for phylogeny finds no significant values despite lambda's small estimated value. All indicator traits are positive on the second component (Fig. 3a) with relative testis size and canine dimorphism having strong loadings, whereas the first component is largely influenced by body mass. Clustering analysis suggests four groups, with motif counts clustered by quadrant and with *c3* and *p21* clustering together. Five motif counts (*psa3*, *hklk*, *hk2*, *cARE*, and *pem2*), three of which significantly correlate with an indicator trait, have positive loadings on both axes (Fig. 3b). Four motif counts (*mvdp*, *psa1*, *slp2*, and *cSRE*) generally have negative loadings on both axes. Species are divided such that Cercopithicoidae is separated from the Hominoidea and Platyrrhini along the second axis. In general, the first axis separates species by the amount of polygyny or precopulatory sexual selection and the second axis separates species by the amount of polygynandry or postcopulatory selection. The results of the multidimensional analysis are in agreement with the previous analyses, making the CCorA (Fig. 3) plot a good visual representation of the data. Motif counts with negative loadings on both axes have negative correlations with all three indicator traits except for *psa1*, which has low correlation values and poor model fits. Three of these motif counts also have a negative loading on the *PC ARE*. Furthermore, motif count positively correlates with testis mass and canine dimorphism and in the CCorA most motif counts are positive on the second axis that is strongly influenced by both of those indicator traits. All five motif counts with positive loadings on both axes also positively correlate with all three indicator traits.
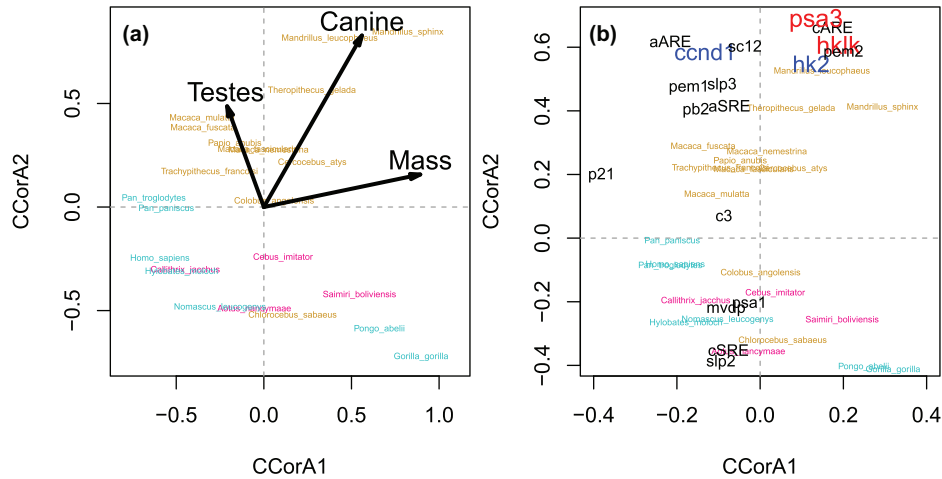
Figure 3. Canonical Correlation Analysis for genome-wide ARE motif counts and indicators of sexual selection in primates. ARE motifs with a significant correlation with one trait in the GLS model are shown in blue and those with two traits are shown in red. Species are colored according to their taxonomic group: fuchsia (Platyrrhini), teal (Hominoidea), and goldenrod (Cercopithecoidea). Panels a and b show the first two canonical variates generated in the CCorA. Scores (correlation between data point and the canonical variate) for indicator traits (a) and ARE motifs (b) are indicated by position on the graph with loadings for species plotted on both panels. For both panels, larger absolute values indicate stronger correlations to the canonical variate axis.

Notably, counts of *mvdp* have repeated negative loading across analyses and are significantly underrepresented in the genes with proximal AREs from the prior analysis of the five species. Both *p21* and *c3* have the lowest ranges of any motif (i.e., max value <70). Combined with the result of overrepresentation of *p21* as a proximal ARE in the expression dataset, our observations suggest that these motifs may have an effect but their low rate of occurrence makes these effects hard to detect when compared to other motifs (typical mean number of occurrences in the genome is ~500).

TRENDS IN MOTIF PATTERNS

Using a haplotype network to show similarities between motifs, we find that the *cARE* is at the center of the network from which the various motifs diverge (Fig. 4). The motifs with genome-wide counts that significantly correlate with indicator traits are not directly adjacent to the *cARE* and are generally dispersed from each other in the network except for *psa3* and *hk2*. No single base-pair position is conserved across the significant motifs (Table 1) except for base pairs that are conserved across nearly all motifs (positions –5, –4, –3, –2). Motifs that are more specific to the androgen receptor are generally closer together on the network and only one (*ccnd1*) of the six motifs is correlated with an indicator trait. Considering motifs that had positive (*cARE* and *pem2*) or negative (*slp2*, *psa1*, *mvdp*, *cSRE*) loadings on the multivariate analysis did not reveal any new noticeable pattern. All of these findings considered together suggest a lack of any consistent evolutionary pressure on specific motif patterns.

Discussion

Our results reveal two general trends in primates. First, AREs are implicated in male-biased, androgen-responsive genes and the gain or loss of proximal AREs over evolutionary time can affect expression values and sex bias of these genes. Second, increasing numbers of genome-wide AREs, along with an increase in sexual dimorphism in trait values, are indicative of sexual selection. Both trends can point to particular motifs that may have a significant effect on expression patterns or have genome-wide counts significantly increase along with indicator traits. No pattern of motif nucleotide sequence, including those with high androgen receptor specificity, predictably influences correlations between traits and genome-wide counts. When considered as a whole, the majority of AREs and the linear combination of AREs do increase in number along with sexual selection and gene expression. We conclude that individual motif sequences may not be as important to sexual selection as the overall trend in the number of motifs, which could allow for stronger sex-biased regulation.

We investigated whether genome-wide ARE motif counts are reliable indicators of sexual selection by examining motif counts as a function of other indicators of sexual selection. The two traits we examined for precopulatory sexual selection are body mass dimorphism and canine height dimorphism, as both have been implicated in sexual selection in primates (Plavcan and van Schaik, 1997, Mitani et al., 1996, Plavcan and van Schaik, 1992, Leutenegger and Kelly, 1977). Body mass dimorphism and canine height dimorphism are strongly correlated ($r = 0.641$, $P = 0.001$, $R^2 = 0.395$, $\lambda = 0.830$), yet canine height has more
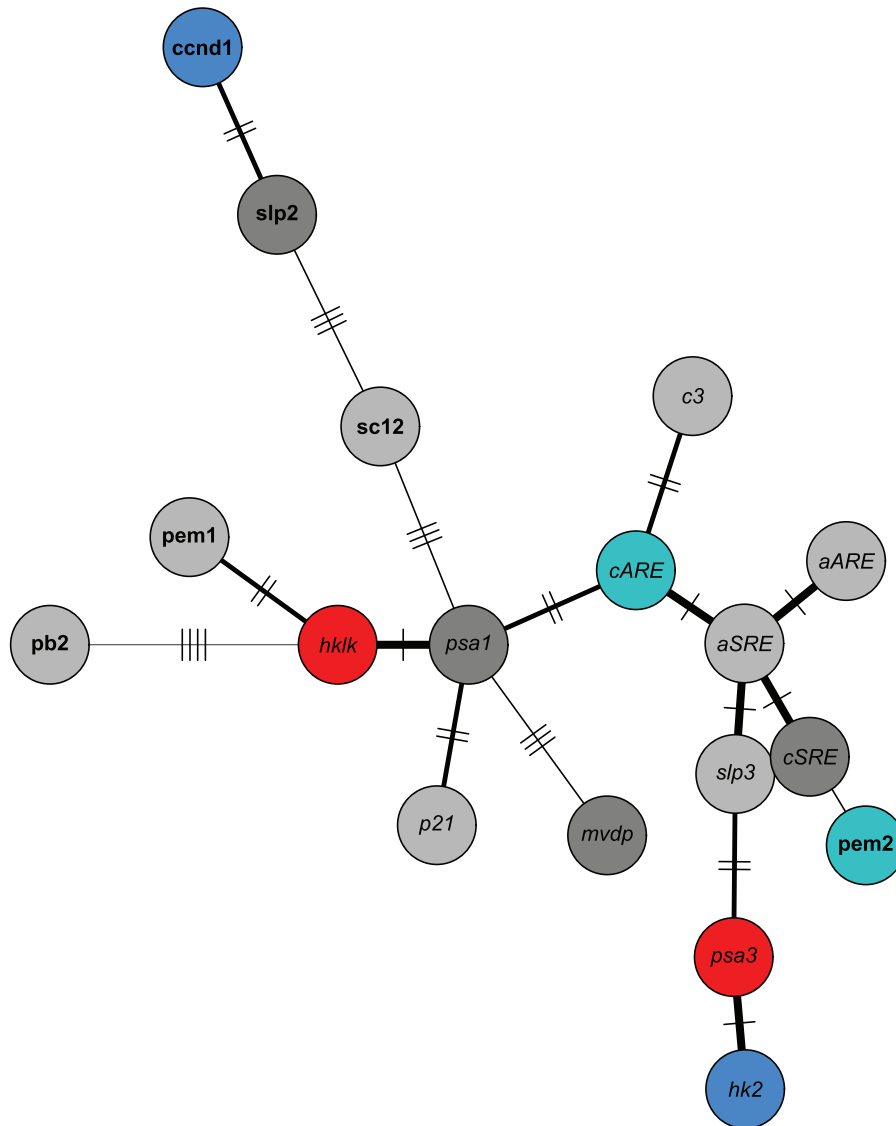
Figure 4. Haplotype network of ARE motifs. Solid lines represent direct connections with line thickness and tick marks corresponding to number of substitutions (thickest = 1, thinnest = 4). Significant correlations with one indicator trait determined from the GLS models are shown in blue and two significant correlations are shown in red. Results from multidimensional analysis (Fig. 2) show additional positively loaded (teal) and negatively loaded (dark gray) motifs. Motifs with known to have high specificity for androgen-receptor binding are in bold.

significantly correlated motifs and more positively correlated motifs. Looking at the multidimensional analyses, body mass dimorphism has little influence on the distribution of ARE motif counts. Instead, canine height dimorphism has a large influence on the distribution of ARE motif counts. If both body size dimorphism and canine size dimorphism are indicators of sexual selection and are correlated, why does genome-wide ARE motif count have a strong signal in one (canine) but no discernable signal in the other (body size)? It is important to note that these traits are approximations for sexual selection and carry assumptions that may not truly reflect the direct influence of the trait on reproductive success (Kappeler and Van Schaik, 2004,

Henshaw et al., 2016, Jones, 2009). Some evidence suggests that body size dimorphism might be more susceptible to factors beyond sexual selection and canine size is a better indicator of sexual selection (Thorén et al., 2006, Leutenegger and Kelly, 1977), although other results call this conclusion into question (Plavcan et al., 1995, Matsuda et al., 2020). Our results cannot settle which trait is a better indicator of sexual selection, but they do show a strong correlation with canine size dimorphism. If canine height is a better predictor of sexual selection than body mass, then the results found here support the hypothesis that greater precopulatory sexual selection will lead to an increase in genome-wide AREs.

To test whether genome-wide ARE motif count is a good indicator of postcopulatory sexual selection, we used relative testis size as our sexually selected trait. Testis size had fewer significant motifs than canine size dimorphism and a weaker effect in the multidimensional analyses but slightly more positive correlations with ARE motifs overall. In ours and other datasets, relative testis mass is not correlated with either body mass ($P = 0.253$) or canine height ($P = 0.981$) sexual dimorphism, suggesting independent effects of pre- and postcopulatory selection (Lupold et al., 2019). Interestingly, the *hklk* and *psa3* motifs, which are affected by canine height and body mass, respectively, are also significantly affected by testis mass. Further, the multidimensional analysis suggests that the direction of testis effects tends to be the same for canine size dimorphism. These results suggest that testis mass does correlate with genome-wide motif number and that precopulatory and postcopulatory sexual selections affect the genome similarly with respect to the evolution of AREs.

A lack of a clear pattern in nucleotide substitutions or androgen specificity in significantly correlated motifs suggests that the precise composition of individual motifs may not be as important as the number of motifs present. Nonetheless, the canonical ARE motif represents the baseline from which the significant motifs are derived and follows the general pattern of increasing in frequency with the evolution of sexually dimorphic traits (Table 1; Fig. 2). Selection and drift likely play roles in further departures from the canonical ARE motif, with selection acting to increase or decrease specificity when necessary and genetic drift primarily affecting neutral or weakly selected substitutions. The motifs that show significant patterns in primates may be a result of conserved genomic regions that are under stronger selection and have had duplications or transpositions in their evolutionary history, thus increasing the number of local motifs. It is likely that a similar study in another taxonomic group will not find the same motifs as significant but would find a similar pattern of an increasing number of motifs correlating with stronger sexual selection.

We suggest a vital role for *cis*-regulatory elements, in particular AREs, in sexual selection as they can help resolve intralocus conflict by using sex-biased signals to create sex-biased gene expression that can lead to sexual dimorphism (Wittman et al., 2021, Cox et al., 2017). Importantly, there are multiple mechanisms for sex-biased expression (Mank, 2017, Ellegren and Parsch, 2007, Wright et al., 2018) and increasing genome-wide counts alone cannot explain the entirety of a genomic response to sexual selection. When androgen regulation does occur, responsive genes can be both *cis*- and *trans*-regulated. When considering the diverse mechanisms for sex bias and *trans*-regulation of androgens, it is no surprise that male-biased genes in humans did not have a significant overrepresentation of proximal AREs. Male-biased genes did skew to the top of the distribution, indicating some effect of AREs. Given the widespread genomic differences in sex-biased trait expression (van der Bijl and Mank, 2021), it is reasonable to propose that some portion of these differences are controlled by AREs in vertebrate systems, as androgens are known to regulate sex-biased phenotypes. Indeed, for genes with a known androgenic response, there is an incredibly apparent overrepresentation of proximal AREs. Not every gene has a proximal ARE, just 15% have one, which is expected given upstream *trans* effects, the possibility of long-range regulation, and undetected AREs. More AREs were indeed present than observed throughout all of our analyses, but our choice to focus on exact matching and a subset of AREs prevents their detection. We expect that our choice represents a subset of AREs and the larger patterns will hold with more AREs, even though individual genes analyzed may change in proximal ARE count. Our results show that AREs do play some role in sex-biased expression and when the genes in question are androgen regulated, they play an incredibly important role.

In our analysis of gain and loss of AREs on gene expression across species, we find a large share of genes that could have correlations with proximal AREs do indeed exhibit such correlations. By filtering genes by male expression levels and sex bias, we are focusing on genes that are likely targets of androgen regulation, which we further ensure by including only samples with at least one species with a proximal ARE. By comparing male RPKM values and sex bias, we can see that the presence of an ARE does not automatically guarantee a sex-biased response. Although caution should be taken due to the small sampling of both species and expression levels, that about one in three genes chosen had a significant correlation between expression or sex bias with proximal AREs is a striking result.

The findings of overrepresentation of proximal AREs in androgen-responsive genes and the effects of gain or loss of proximal AREs on gene expression and sex bias demonstrate the influential roles AREs have in sex-biased traits. When considering genome-wide counts, we view an increase in AREs as both creating new locations for *cis*-regulation and enhancing responses in already existing androgen-responsive regions. Not every discovered ARE is necessarily active; nucleotides flanking the half-sites and inside the spacer may influence binding (Nelson et al., 1999), and the presence of the FOXA1 promoter may influence the expression of androgen-induced genes (Jin et al., 2014, Kregel et al., 2020, Sahu et al., 2011). Further, a gene with androgen regulation may not necessarily be sex biased or under sexual selection. Although these factors individually may affect androgen binding, the goal of our study was to investigate global patterns of the ARE motif in response to sexual selection. A pattern is apparent based on these traits and expression changes, particularly relative testis size and canine sexual dimorphism, which are indicators of strong sexual selection.

ARE plays a crucial role in the development of male-biased traits when sexual selection acts strongly on males. Our results show that there is a corresponding increase in genome-wide AREs with increasing sexual selection in primates. As more fully sequenced genomes become available, the robustness of these findings can be further explored in primates and across multiple vertebrate systems. Although some motifs show multiple significant patterns, there is no indication that androgen receptor specificity or nucleotide sequence (as opposed to the number of AREs) is the main target of selection, suggesting that different clades might favor alternate motifs depending on their evolutionary histories. Ultimately, an analysis of additional genomes across a wider array of organisms will better elucidate the evolutionary response of genome-wide ARE motif number to an increase in sexual selection intensity. Given these findings, there is a clear need to further investigate the role that AREs and other noncoding DNA elements play in evolutionary responses to natural and sexual selection.

AUTHOR CONTRIBUTIONS

APA conceived the project, collected the data, performed the analysis, designed the figures, and wrote the manuscript. AGJ conceived the project and wrote the manuscript.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

DATA ARCHIVING

All relevant R code and data tables can be accessed in the Supporting Information and at https://github.com/AndersonDrew/Primate_ARE.

REFERENCES

Abouheif, E., & Fairbairn, D.J. (1997) A comparative analysis of allometry for sexual size dimorphism: assessing Rensch's rule. *The American Naturalist*, Mar 1;149(3), 540–62.

Adler, A.J., Danielsen, M., & Robins, D.M. (1992) Androgen-specific gene activation via a consensus glucocorticoid response element is determined by interaction with nonreceptor factors. *Proc. Natl Acad. Sci. USA*, 89(24), 11660–3.

Akaike, H. (1974) A new look at the statistical model identification. *IEEE Trans. Autom. Control*,. 19(6), 716–23.

Anderson, A.P., Flanagan, S.P., Rose, E., & Jones, A.G. (2020) The estrogen-responsive transcriptome of female secondary sexual traits in the Gulf pipefish. *J. Hered.*, 111(3), 294-306.

Anderson, A.P., & Jones, A.G. (2019) erefinder: genome-wide detection of oestrogen response elements. *Mol. Ecol. Resour.*, 19(5), 1366–73.

Anderson, M.J., & Dixson, A.F. (2002) Motility and the midpiece in primates. *Nature*,. Apr 1;416(6880), 496–496.

Arnold, C., Matthews, L.J., & Nunn, C.L. (2010) The 10kTrees website: a new online resource for primate phylogeny. *Evol Anthropol Issues News Rev*,. 19(3), 114–8.

Barbulescu, K., Geserick, C., Schüttke, I., Schleuning, W.-.D., & Haendler, B. (2001) New androgen response elements in the murine pem promoter mediate selective transactivation. *Mol. Endocrinol.*, 15(10), 1803–16.

Bartos, L., Bubenik, G.A., & Kuzmova, E. (2012) Endocrine relationships between rank-related behavior and antler growth in deer. *Front. Biosci.*, 4, 1111–26.

Bhattacharya, I., Basu, S., Pradhan, B.S., Sarkar, H., Nagarajan, P., & Majumdar, S.S. (2019) Testosterone augments FSH signaling by upregulating the expression and activity of FSH-Receptor in Pubertal Primate Sertoli cells. *Mol. Cell. Endocrinol.*, Feb 15;482, 70–80.

Bonduriansky, R., & Chenoweth, S.F. (2009) Intralocus sexual conflict. *Trends Ecol. Evol.*, May 1;24(5), 280–8.

Brawand, D., Soumillon, M., Necsulea, A., Julien, P., Csárdi, G., Harrigan, P., Weier, M., Liechti, A., Aximu-Petri, A., Kircher, M., et al (2011) The evolution of gene expression levels in mammalian organs. *Nature*,. Oct 1;478(7369), 343–8.

Carroll, J.S., Liu, X.S., Brodsky, A.S., Li, W., Meyer, C.A., Szary, A.J., Eeckhoute, J., Shao, W., Hestermann, E.V., Geistlinger, T.R., et al (2005) Chromosome-wide mapping of estrogen receptor binding reveals long-range regulation requiring the forkhead protein FoxA1. *Cell*,. Jul 15;122(1), 33–43.

Claessens, F., Alen, P., Devos, A., Peeters, B., Verhoeven, G., & Rombauts, W. (1996) The androgen-specific probasin response element 2 interacts differentially with androgen and glucocorticoid receptors. *J. Biol. Chem.*, 271(32), 19013–6.

Claessens, F., Celis, L., Peeters, B., Heyns, W., Verhoeven, G., & Rombauts, W. (1989) Functional characterization of an androgen response element in the first intron of the C3 (1) gene of prostatic binding protein. *Biochem. Biophys. Res. Commun.*, 164(2), 833–40.

Claessens, F., Joniau, S., & Helsen, C. (2017) Comparing the rules of engagement of androgen and glucocorticoid receptors. *Cell. Mol. Life Sci.*, 74(12), 2217–28.

Cleutjens, K.B., vanEekelen, C.C., van derKorput, H.A., Brinkmann, A.O., & Trapman, J. (1996) Two androgen response regions cooperate in steroid hormone regulated activity of the prostate-specific antigen promoter. *J. Biol. Chem.*, 271(11), 6379–88.

Cleutjens, K., van derKorput, H., vanEekelen, C., vanRooij, H.C.J., Faber, P.W., & Trapman, J. (1997) An androgen response element in a far upstream enhancer region is essential for high, androgen-regulated activity of the prostate-specific antigen promoter. *Mol. Endocrinol.*, Feb 1;11(2), 148–61.

Clutton-Brock, T.H. (1985) Size, sexual dimorphism, and polygyny in primates. Pp. 51–60*in*W. L.Jungers, ed. Size and scaling in primate biology. Springer, Berlin.

Connallon, T., & Clark, A.G. (2014) Evolutionary inevitability of sexual antagonism. *Proc R Soc B Biol Sci*,. Feb 7;281(1776), 20132123.

Cox, R.M., Cox, C.L., McGlothlin, J.W., Card, D.C., Andrew, A.L., & Castoe, T.A. (2017) Hormonally mediated increases in sex-biased gene expression accompany the breakdown of between-sex genetic correlations in a sexually dimorphic lizard. *Am. Nat.*, Mar 1;189(3), 315–32.

Dahl, J.F., Gould, K.G., & Nadler, R.D. (1993) Testicle size of orang-utans in relation to body size. *Am. J. Phys. Anthropol.*, 90(2), 229–36.

Denayer, S., Helsen, C., Thorrez, L., Haelens, A., & Claessens, F. (2010) The rules of DNA recognition by the androgen receptor. *Mol. Endocrinol.*, May 1;24(5), 898–913.

Dines, J.P., Otárola-Castillo, E., Ralph, P., Alas, J., Daley, T., Smith, A.D., & Dean, M.D. (2014) Sexual selection targets cetacean pelvic bones. *Evolution*, 68, 3296–3306.

Ellegren, H., & Parsch, J. (2007) The evolution of sex-biased genes and sex-biased gene expression. *Nat. Rev. Genet.*, Sep 1;8(9), 689–98.

Fabre, S., Manin, M., Pailhoux, E., Veyssière, G., & Jean, C. (1994) Identification of a functional androgen response element in the promoter of the gene for the androgen-regulated aldose reductase-like protein specific to the mouse vas deferens. *J. Biol. Chem.*, 269(8), 5857–64.

Fitzpatrick, J.L., Almbro, M., Gonzalez-Voyer, A., Hamada, S., Pennington, C., Scanlan, J., & Kolm, N. (2012) Sexual selection uncouples the evolution of brain and body size in pinnipeds. *J. Evol. Biol.*, 25, 1321–1330.

Fitzpatrick, J.L., Almbro, M., Gonzalez-Voyer, A., Kolm, N., & Simmons, L.W. (2012) Male contest competition and the coevolution of weaponry and testes in pinnipeds. *Evolution.*, Nov 1;66(11), 3595–604.

Geserick, C., Meyer, H.-.A., Barbulescu, K., & Haendler, B. (2003) Differential modulation of androgen receptor action by deoxyribonucleic acid response elements. *Mol. Endocrinol.*, Sep 1;17(9), 1738–50.

Geserick, C., Meyer, H.-.A., & Haendler, B. (2005) The role of DNA response elements as allosteric modulators of steroid receptor function. *Mol. Cell. Endocrinol.*, May 31;236(1), 1–7.

Ghosal, R., & Sorensen, P.W. (2016) Male-typical courtship, spawning behavior, and olfactory sensitivity are induced to different extents by androgens in the goldfish suggesting they are controlled by different neuroendocrine mechanisms. *Gen. Comp. Endocrinol.*, Jun 1;232, 160–73.

Guo, S., Zhou, Y., Zeng, P., Xu, G., Wang, G., & Cui, Q. (2018) Identification and analysis of the human sex-biased genes. *Brief Bioinform*,. Mar 1;19(2), 188–98.

Hahn, M.W. (2007) Detecting natural selection on cis-regulatory DNA. *Genetica*,. 129(1), 7–18.

Harcourt, A.H., Purvis, A., & Liles, L. (1995) Sperm competition: mating system, not breeding season, affects testes size of primates. *Funct Ecol*,. 9(3), 468–76.

Harold, H. (1936) Relations between two sets of variates. *Biometrika*,. 28(3/4), 321–77.

Harrison, P.W., Wright, A.E., Zimmer, F., Dean, R., Montgomery, S.H., Pointer, M.A., Mank, J.E. (2015) Sexual selection drives evolution and rapid turnover of male gene expression. *Proc. Natl Acad. Sci.*, Apr 7;112(14), 4393.

Henshaw, J.M., Kahn, A.T., & Fritzsche, K. (2016) A rigorous comparison of sexual selection indexes via simulations of diverse mating systems. *Proc. Natl Acad. Sci. USA*, Jan 19;113(3), E300–E308.

Hill, W.O., & Kanagasuntheram, R. (1959) The male reproductive organs in certain gibbons (Hylobatidae). *Am. J. Phys. Anthropol.*, 17(3), 227–41.

Jiménez-Delgado, S., Pascual-Anaya, J., & Garcia-Fernàndez, J. (2009) Implications of duplicated cis-regulatory elements in the evolution of metazoans: the DDI model or how simplicity begets novelty. *Brief Funct Genomics*,. Jul 1;8(4), 266–75.

Jin, H.-.J., Zhao, J.C., Wu, L., Kim, J., & Yu, J. (2014) Cooperativity and equilibrium with FOXA1 define the androgen receptor transcriptional program. *Nat. Commun.*, May 30;5(1), 3972.

Jones, A.G. (2009) On the opportunity for sexual selection, the bateman gradient and the maximum intensity of sexual selection. *Evolution.*, 63(7), 1673–84.

Kahrl, A.F., Cox, C.L., & Cox, R.M. (2016) Correlated evolution between targets of pre- and postcopulatory sexual selection across squamate reptiles. *Ecol Evol*,. Sep 1;6(18), 6452–9.

Kappeler, P.M., & Van Schaik, C.P. (2004) Sexual selection in primates: new and comparative perspectives. Cambridge Univ. Press, Cambridge, U.K.

Kirkpatrick, M. (2017) The evolution of genome structure by natural and sexual selection. *J. Hered.*, 108, 3–11.

Kregel, S., Bagamasbad, P., He, S., LaPensee, E., Raji, Y., Brogley, M., Chinnaiyan, A., Cieslik, M., & Robins, D.M. (2020) Differential modulation of the androgen receptor for prostate cancer therapy depends on the DNA response element. *Nucleic. Acids. Res.*, May 21;48(9), 4741–55.

Lande, R. (1980) Sexual Dimorphism, Sexual Selection, and Adaptation in Polygenic Characters. *Evolution.*, 34(2), 292–305.

Lanzino, M., Sisci, D., Morelli, C., Garofalo, C., Catalano, S., Casaburi, I., Capparelli, C., Giordano, C., Giordano, F., Maggiolini, M., et al (2010) Inhibition of cyclin D1 expression by androgen receptor in breast cancer cells—identification of a novel androgen response element. *Nucleic. Acids. Res.*, 38(16), 5351–65.

Laudet, V., Hänni, C., Coll, J., Catzeflis, F., & Stéhelin, D. (1992) Evolution of the nuclear receptor gene superfamily. *EMBO J.*, 11(3), 1003–13.

Leutenegger, W., & Kelly, J.T. (1977) Relationship of sexual dimorphism in canine size and body size to social, behavioral, and ecological correlates in anthropoid primates. *Primates*,. 18(1), 117–36.

Lindsay, W.R., Barron, D.G., Webster, M.S., & Schwabl, H. (2016) Testosterone activates sexual dimorphism including male-typical carotenoid but not melanin plumage pigmentation in a female bird. *J. Exp. Biol.*, Oct 1;219(19), 3091–9.

Lu, S., Liu, M., Epner, D.E., Tsai, S.Y., & Tsai, M.-.J. (1999) Androgen regulation of the cyclin-dependent kinase inhibitor p21 gene through an androgen response element in the proximal promoter. *Mol. Endocrinol.*, 13(3), 376–84.

Lüpold, S., Simmons, L.W., & Grueter, C.C. (2019) Sexual ornaments but not weapons trade off against testes size in primates. *Proc. R. Soc. B Biol. Sci.*, 286(1900), 20182542.

Malika, C., Ghazzali, N., Boiteau, V., & Niknafs, A. (2014) NbClust: an R package for determining the relevant number of clusters in a data set. *J Stat Softw*,. 61, 1–36.

Mank, J.E. (2017) Population genetics of sexual conflict in the genomic era. *Nat. Rev. Genet.*, 18, 721–730.

Matsuda, I., Stark, D.J., Saldivar, D.A.R., Tuuga, A., Nathan, S., Goossens, B., vanSchaik, CP., & Koda, H. (2020) Large male proboscis monkeys have larger noses but smaller canines. *Commun Biol*,. Sep 21;3(1), 522.

Matthews, B.J., Melia, T., & Waxman, D.J. (2021) Harnessing natural variation to identify cis regulators of sex-biased gene expression in a multistrain mouse liver model. *PLos Genet.*, Nov 9;17(11), e1009588.

Mitani, J.C., Gros-Louis, J., & Richards, A.F. (1996) Sexual dimorphism, the operational sex ratio, and the intensity of male competition in polygynous primates. *Am. Nat.*, Jun 1;147(6), 966–80.

Mitchell, S.H., Murtha, P.E., Zhang, S., Zhu, W., & Young, C.Y. (2000) An androgen response element mediates LNCaP cell dependent androgen induction of the hK2 gene. *Mol. Cell. Endocrinol.*, 168(1–2), 89–99.

Møller, A.P. (1988) Ejaculate quality, testes size and sperm competition in primates. *J. Hum. Evol.*, Aug 1;17(5), 479–88.

Morris, J.S., & Carrier, D.R. (2016) Sexual selection on skeletal shape in Carnivora. *Evolution.*, Apr 1;70(4), 767–80.

Murtha, P., Tindall, D.J., & Young, C.Y. (1993) Androgen induction of a human prostate-specific kallikrein, hKLK2: characterization of an androgen response element in the 5'promoter region of the gene. *Biochemistry*,. 32(25), 6459–64.

Nelson, C.C., Hendy, S.C., Shukin, R.J., Cheng, H., Bruchovsky, N., Koop, B.F., & Rennie, P.S. (1999) Determinants of DNA sequence specificity of the androgen, progesterone, and glucocorticoid receptors: evidence for differential steroid receptor response elements. *Mol. Endocrinol.*, Dec 1;13(12), 2090–107.

Nelson, P.S., Clegg, N., Arnold, H., Ferguson, C., Bonham, M., White, J., Hood, L., & Lin, B. (2002) The program of androgen-responsive genes in neoplastic prostate epithelium. *Proc Natl Acad. Sci.*, Sep 3;99(18), 11890.

Oliva, M., Manuel, M.-.A., Sarah, K.-.H., Wucher, V., Gewirtz, A.D.H., Cotter, D.J., Parsana, P., Kasela, S., Balliu, B., Viñuela, A., et al (2020) The impact of sex on gene expression across human tissues. *Science*,. Sep 11;369(6509), eaba3066.

Pagès, H., Aboyoun, P., Gentleman, R., & DebRoy, S. (2019) Biostrings: efficient manipulation of biological strings. R package version 2.0.

Paradis, E. (2010) pegas: an R package for population genetics with an integrated–modular approach. *Bioinformatics*,. 26(3), 419–20.

Pennell, T.M., & Morrow, E.H. (2013) Two sexes, one genome: the evolutionary dynamics of intralocus sexual conflict. *Ecol Evol*,. Jun 1;3(6), 1819–34.

Plavcan, J.M., & Ruff, C.B. (2008) Canine size, shape, and bending strength in primates and carnivores. *Am. J. Phys. Anthropol*,. 136(1), 65–84.

Plavcan, J.M., & vanSchaik, C.P. (1992) Intrasexual competition and canine dimorphism in anthropoid primates. *Am. J. Phys. Anthropol.*, Apr 1;87(4), 461–77.

Plavcan, J.M., & vanSchaik, C.P. (1997) Intrasexual competition and body weight dimorphism in anthropoid primates. *Am. J. Phys. Anthropol.*, May 1;103(1), 37–68.

Plavcan, J.M., vanSchaik, C.P., & Kappeler, P.M. (1995) Competition, coalitions and canine size in primates. *J. Hum. Evol.*, Mar 1;28(3), 245–76.

Pointer, M.A., Harrison, P.W., Wright, A.E., & Mank, J.E. (2013) Masculinization of Gene Expression Is Associated with Exaggeration of Male Sexual Dimorphism. *PLos Genet.*, Aug 15;9(8), e1003697.

Rensch, B. (1950) Die Abhängigkeit der relativen Sexualdifferenz von der Körpergrösse. *Bonn Zool Beitr*,. 1, 58–69.

Revell, L.J. (2012) phytools: an R package for phylogenetic comparative biology (and other things). *Methods Ecol. Evol.*, 3(2), 217–23.

Riegman, P.H.J., Vlietstra, R.J., Van der Korput, J., Brinkmann, A.O., & Trapman, J. (1991) The promoter of the prostate-specific antigen gene contains a functional androgen responsive element. *Mol. Endocrinol.*, 5(12), 1921–30.

Roche, P.J., Hoare, S.A., & Parker, M.G. (1992) A consensus DNA-binding site for the androgen receptor. *Mol. Endocrinol.*, Dec 1;6(12), 2229–35.

Romanuik, T.L., Wang, G., Holt, R.A., Jones, S.J., Marra, M.A., & Sadar, M.D. (2009) Identification of novel androgen-responsive genes by sequencing of LongSAGE libraries. *BMC Genomics*,. Oct 15;10(1), 476.

Sahu, B., Laakso, M., Ovaska, K., Mirtti, T., Lundin, J., Rannikko, A., Sankila, A., Turunen, J.-.P., Lundin, M., Konsti, J. et al (2011) Dual role of FoxA1 in androgen receptor binding to chromatin, androgen signalling and prostate cancer. *EMBO J.*, Oct 5;30(19), 3962–76.

Sahu, B., Pihlajamaa, P., Dubois, V., Kerkhofs, S., Claessens, F., & Jänne, O.A. (2014) Androgen receptor uses relaxed response element stringency for selective chromatin binding and transcriptional regulation in vivo. *Nucleic. Acids. Res.*, Apr 1;42(7), 4230–40.

Schauwaers, K., De Gendt, K., Saunders, P.T.K., Atanassova, N., Haelens, A., Callewaert, L., Moehren, U., Swinnen, J.V., Verhoeven, G., Verrijdt, G., et al (2007) Loss of androgen receptor binding to selective androgen response elements causes a reproductive phenotype in a knockin mouse model. *Proc. Natl Acad. Sci. USA*, Mar 20;104(12), 4961.

Scrucca, L., Fop, M., Murphy, T.B., & Raftery, A.E. (2016) mclust 5: clustering, classification and density estimation using Gaussian finite mixture models. *R J*,. 8(1), 289.

Simão, F.A., Waterhouse, R.M., Ioannidis, P., Kriventseva, E.V., & Zdobnov, E.M. (2015) BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics*,. 31(19), 3210–2.

Smith, J.D., McManus, K.F., & Fraser, H.B. (2013) A Novel Test for Selection on cis-Regulatory Elements Reveals Positive and Negative Selection Acting on Mammalian Transcriptional Enhancers. *Mol. Biol. Evol.*, Nov 1;30(11), 2509–18.

Stewart, A.D., Pischedda, A., & Rice, W.R. (2010) Resolving intralocus sexual conflict: genetic mechanisms and time frame. *J. Hered.*, Mar 1;101(suppl_1), S94–9.

Thorén, S., Lindenfors, P., & Kappeler, P.M. (2006) Phylogenetic analyses of dimorphism in primates: evidence for stronger selection on canine size than on body size. *Am. J. Phys. Anthropol.*, May 1;130(1), 50–9.

Tsai, M.-.J., & O'Malley, B.W. (1994) Molecular mechanisms of action of steroid/thyroid receptor superfamily members. *Annu. Rev. Biochem.*, Jun 1;63(1), 451–86.

van derBijl, W., & Mank, J.E. (2021) Widespread cryptic variation in genetic architecture between the sexes. *Evol Lett*,. 5(4), 359–369

Verrijdt, G., Haelens, A., & Claessens, F. (2003) Selective DNA recognition by the androgen receptor as a mechanism for hormone-specific regulation of gene expression. *Mol. Genet. Metab.*, Mar 1;78(3), 175–85.

Verrijdt, G., Schoenmakers, E., Alen, P., Haelens, A., Peeters, B., Rombauts, W., & Claessens, F. (1999) Androgen specificity of a response unit upstream of the human secretory component gene is mediated by differential receptor binding to an essential androgen response element. *Mol. Endocrinol.*, 13(9), 1558–70.

Verrijdt, G., Schoenmakers, E., Haelens, A., Peeters, B., Verhoeven, G., Rombauts, W., & Claessens, F. (2000) Change of specificity mutations in androgen-selective enhancers evidence for a role of differential DNA binding by the androgen receptor. *J. Biol. Chem.*, 275(16), 12298–305.

Wasserman, W.W., & Sandelin, A. (2004) Applied bioinformatics for the identification of regulatory elements. *Nat. Rev. Genet.*, Apr 1;5(4), 276–87.

Welboren, W.-.J., Van Driel, M.A., Janssen-Megens, E.M., Van Heeringen, S.J., Sweep, F.C., Span, P.N., & Stunnenberg, H.G. (2009) ChIP-Seq of ERα and RNA polymerase II defines genes differentially responding to ligands. *EMBO J.*, 28(10), 1418–28.

Wilkinson, G.S., Breden, F., Mank, J.E., Ritchie, M.G., Higginson, A.D., Radwan, J., Jaquiery, J., Salzburger, W., Arriero, E., Barribeau, S.M., et al (2015) The locus of sexual selection: moving sexual selection studies into the post-genomics era. *J. Evol. Biol.*, 28, 739–755.

Wilson, S., Qi, J., & Filipp, F.V. (2016) Refinement of the androgen response element based on ChIP-Seq in androgen-insensitive and androgen-responsive prostate cancer cell lines. *Sci. Rep.*, Sep 14;6(1), 32611.

Wittkopp, P.J., & Kalay, G. (2012) Cis-regulatory elements: molecular mechanisms and evolutionary processes underlying divergence. *Nat. Rev. Genet.*, Jan 1;13(1), 59–69.

Wittman, T.N., Robinson, C.D., McGlothlin, J.W., & Cox, R.M. (2021) Hormonal pleiotropy structures genetic covariance. *Evol Lett*, 5(4), 397-407

Wray, G.A. (2007) The evolutionary significance of cis-regulatory mutations. *Nat. Rev. Genet.*, Mar 1;8(3), 206–16.

Wright, A.E., Fumagalli, M., Cooney, C.R., Bloch, N.I., Vieira, F.G., Buechel, S.D., Kolm, N., & Mank, J.E. (2018) Male-biased gene expression resolves sexual conflict through the evolution of sex-specific genetic architecture. *Evol Lett*,. 2(2), 52–61.

Zheng, Y., Shao, X., Huang, Y., Shi, L., Chen, B., Wang, X., Yang, H., Chen, Z., & Zhang, X. (2016) Role of estrogen receptor in breast cancer cell gene expression. *Mol Med Rep*,. 13(5), 4046–50.

Associate Editor: T. Connallon
Handling Editor: A. G. McAdam

Supporting Information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Supporting Information
Supporting Information
Supporting Information
Supporting Information
Supporting Information
Supporting Information
Supporting Information
Supporting Information