bioRxiv preprint doi: https://doi.org/10.1101/125799; this version posted April 8, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Comparative genomic evidence for self-domestication in *Homo sapiens*

Constantina Theofanopoulou[1,2,a], Simone Gastaldon[1,a], Thomas O'Rourke[1], Bridget D. Samuels[3], Angela Messner[1], Pedro Tiago Martins[1], Francesco Delogu[4], Saleh Alamri[1], and Cedric Boeckx[1,2,5,b]

[1]Section of General Linguistics, Universitat de Barcelona
[2]Universitat de Barcelona Institute for Complex Systems
[3]Center for Craniofacial Molecular Biology, University of Southern California
[4]Department of Chemistry, Biotechnology and Food Science, Norwegian University of Life Sciences (NMBU)
[5]ICREA
[a]These authors contributed equally to this work
[b]Corresponding author: cedric.boeckx@ub.edu

ABSTRACT

This study identifies and analyzes statistically significant overlaps between selective sweep screens in anatomically modern humans and several domesticated species. The results obtained suggest that (paleo-)genomic data can be exploited to complement the fossil record and support the idea of self-domestication in *Homo sapiens*, a process that likely intensified as our species populated its niche. Our analysis lends support to attempts to capture the "domestication syndrome" in terms of alterations to certain signaling pathways and cell lineages, such as the neural crest.

Introduction

Recent advances in genomics, coupled with other sources of information, offer new opportunities to test long-standing hypotheses about human evolution. Especially in the domain of cognition, the retrieval of ancient DNA could, with the help of well-articulated linking hypotheses connecting genes, brain and cognition, shed light on the emergence of 'cognitive modernity'. It is in this context that we would like to adduce evidence for an old hypothesis about the evolution of our species: self-domestication. As is well-documented[1], several scholars have entertained the idea that anatomically modern humans (AMH) were self-domesticated. More recently, Hare[2] articulated a more solid hypothesis bringing together pieces of strongly suggestive evidence.

It is clear that our species is unusually dependent on its socio-cultural environment to prosper. As Tomasello has stressed on numerous occasions (e.g.,[3]), our instinct to cooperate, possibly facilitated by a decrease in emotional reactivity[4], marks us as 'special', certainly among primates, with the possible exception of bonobos, which have been claimed to be 'self-domesticated'[5]. Our social tolerance and instinct to learn from members of our community are perhaps most obvious in the domain of language acquisition, which has led to claims that a self-domestication process in our species may have gone hand in hand with the emergence of a fully modern 'language-ready' brain[6–8], and the construction and functioning of our cultural niche[1,9]. Apart from this behavioral uniqueness, several anatomical characteristics, especially when contrasted with what we know about Neanderthals (the best-studied example of archaic *Homo*), are reminiscent of what one finds in comparisons of domesticated species and their wild counterparts.

In order to detect the effects of domestication, data must come from comparisons of a domesticated species with "either their direct wild-living ancestor or close relatives if the ancestor is no longer extant"[1]. In the case of AMH, since there is no wild extant counterpart available, the obvious comparanda include our closest living relatives, i.e., the great apes, and extinct species of the genus *Homo* to the extent that relevant data can be extracted from the fossil record. Crucially for the purposes of this paper, we now have high-quality genomes for our closest extinct relatives, allowing for a more focused comparison. These comparisons can shed light on when the self-domestication event was initiated in our human lineage. We contend (*contra*[10]) that self-domestication coincided with the emergence of AMH (*sensu*[11]: specimens sharing a significant number of derived features in the skeleton with extant members of our species), because, as will be discussed below, some of the critical traits associated with the domestication process are defining features of our species' anatomical modernity, although of course it is very likely the case that this self-domestication process intensified as our species expanded geographically and demographically.

Domesticated species display a range of anatomical and behavioral characteristics that set them apart from their wild

bioRxiv preprint doi: https://doi.org/10.1101/125799; this version posted April 8, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

counterparts: depigmentation; floppy, reduced ears; shorter muzzles; curly tails; smaller teeth; smaller cranial capacity (and concomitant brain size reduction); neotenous (juvenile) behavior; reduction of sexual dimorphism (feminization); docility; and more frequent estrous cycles. Of course, not all of these characteristics are found in all domesticates, but many of them are indeed present to some extent in all domesticates[12]. This constellation of features is known as the "domestication syndrome", and has been hypothesized to arise from a mild deficit of neural crest cells[13], to which we return below.

Many of the anatomical changes associated with domestication describe some of the well-known anatomical differences between AMH and Neanderthals (see Figure 1). The two species display different ontogenetic trajectories[14,15] resulting in craniofacial differences that invariably lead to a more 'gracile', 'juvenile' profile in AMH relative to Neanderthals. It is well-established that prognathism is significantly reduced in our species[15,16]. Brow ridges and nasal projections are smaller in AMH than in our most closely related (extinct) relatives[10], as is our cranial capacity[17] and our tooth size[18,19]. This profile is sometimes called 'feminized'[10], and is associated with an overall reduction of sexual dimorphism, which is a trait associated with domestication[5]. The process of 'feminization' (reduction of androgen levels and rise in estrogen levels[10]) is often associated with reduced reactivity of the hypothalamus-pituitary-adrenal axis[20], a physiological trait thought to be critical for domestication[13,21]. Evidence from digit ratio comparisons—a measure of prenatal androgen exposure, with lower second digit:fourth digit ratios pointing to higher prenatal androgen exposure, and therefore potentially a higher proclivity for aggressiveness[22]—further suggests that Neanderthals had higher prenatal androgen exposure than AMH[23]. Additional differences in other traits associated with domestication may exist, but there are either obvious confounding factors (e.g., geography for pigmentation) involved, or the data are subject to more controversial interpretation (e.g., in the case of reproductive cycle changes[24]) than the features reviewed above.

The central claim of this paper is that there is an additional layer of evidence in favor of self-domestication in anatomically modern humans that can be constructed from findings from (paleo-)genomics. We now have detailed information concerning the genomes of several domesticated species and their wild counterparts[25], as well as high-quality data from archaic human genomes[26]. This information offers the opportunity to test for the existence of overlapping regions of putative signatures of selection associated with (self-)domestication. If found to exist, such overlaps would complement the anatomical data briefly reviewed above and suggest that the self-domestication hypothesis is a strong contender to account for key aspects of modern human cognition.

Results

We examined the overlap of gene sets independently claimed to be under positive selection in AMH (when compared with Neanderthal/Denisovan) and several domesticates for which detailed genetic information is available: dog (*Canis familiaris*), cat (*Felis catus*), horse (*Equus caballus*) and taurine cattle (*Bos taurus*). The pool of domesticates chosen yielded a total of 691 genes, and the total AMH pool, 742 genes. The intersection size (i.e., the number of genes that are positively selected or under selective sweep both in AMH and in one or more domesticate) returns a statistically significant number of 41 ($p < 0.01$) (See Table S2 and Figure 2). We confirmed the significance of this result with a Monte Carlo simulation of 1,000,000 trials, in which samples of 691 and 742 genes were randomly selected (with no replacement) from a pool of 19,500. The simulation confirmed that an intersection size greater than or equal to 41 is highly significant ($p = 0.0003$).

This situation contrasted with the modest (statistically insignificant) overlaps between the domesticates and several Great Apes for which selective sweep screens were available: chimpanzee (*Pan t. troglodytes*), orangutan (*Pongo abelii*), and gorilla (*G. g. gorilla*) (Table S4).

Intersections between domesticates (15 genes in total, see Table S1) were tested, revealing a significant overlap between genes selected in the dog and in cattle ($p < 0.01$). Furthermore, tests between AMH and each domesticate showed significant overlaps with the dog ($v = 15$, $p < 0.05$) and with cattle ($v = 9$, $p < 0.01$). Since we pooled data for positive selection and selective sweeps in AMH from different sources, intersection tests were carried out between the domestication pool and the pool of each AMH dataset used in this study. A significant intersection was found with the data in[27] ($p < 0.05$) and with the combined data from[27] and[28] ($p < 0.05$). Intersections between domesticates were also tested, revealing a significant overlap between genes selected in the dog and in cattle ($p < 0.01$).

Though no gene was found to be shared across all domesticated species studied here as well as AMH, this is not necessarily expected—domestication is known to proceed through various routes (see Discussion), and is thus not a uniform affair. However, common pathways can be identified, as can genes that may have contributed to domestication events that we think deserve special attention. Five genes were found to be associated with signals of positive selection in AMH and multiple domesticated species (Table S3): *RNPC3, FAM172A, PLAC8L1, GRIK3* and *BRAF*.

RNPC3 shows evidence of positive selection in the dog, cat, and AMH. *RNPC3* is one of only two genes with more than one putatively causal variant fixed between dogs and wolves (the other is a gene of unknown function)[29]. Mutations in *RNPC3* cause growth hormone deficiencies in humans resulting from pituitary hypoplasia[30,31]. In a similar vein, a gene showing an AMH-specific amino acid change and associated with a strong positive selection signal in AMH and in dogs,

bioRxiv preprint doi: https://doi.org/10.1101/125799; this version posted April 8, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

NCOA6, is a nuclear receptor coactivator that directly binds nuclear receptors and stimulates the transcriptional activities in a hormone-dependent fashion.

FAM172A, selected for in dogs, cattle, and AMH, may perhaps be worthy of note given its position on chromosome 5 neighboring *NR2F1*, which plays a role in regulating neural crest specifier genes and has undergone selection in AMHs[32,33]; The functionally related nuclear receptor *NR2F2*, involved in regulating embryonic stem-cell differentiation[34], and implicated in neural crest development, has been selected for in the domesticated fox[35].

PLAC8L1 is selected for in more than one domesticate species (cat, cattle) as well as in AMH, but there is only sparse evidence concerning its function. An autistic patient has been noted as having a microdeletion at chromosome 5q32, a location which includes *PLAC8L1*[36].

The two remaining genes in Table S3, *BRAF* and *GRIK3*, deserve special attention.

ERK pathway

BRAF, selected for in the cat, horse, and AMH, is an important member of the ERK/MAPK signaling pathway, which has been shown to play a key role in synaptic plasticity, memory, and learning[37], and which, when disrupted, can lead to a broad range of syndromes comprising craniofacial defects and cognitive deficits[38]. *BRAF* is upstream of *ERK2*, which plays a critical role in neural crest development[39] and regulates neuronal gene expression in both the neocortex and hippocampus[37]. Both *BRAF* and *ERK2* inactivation can bring about syndromic symptoms by disrupting neural crest development[39]. *BRAF* is implicated in Noonan, Leopard, and Cardiofaciocutaneous syndromes, typical symptoms of which include prominent forehead, bitemporal narrowing, hypertelorism, short stature, among other skeletal, cardiac, and craniofacial anomalies, frequently accompanied by moderate to severe mental retardation[40,41]. *BRAF*, interacts with other domestication-related genes, including *YWHAH* (dog), *PPP2CA* (a neural crest-related gene, selected for in the horse), and *HER4/ERBB4*, another neural crest gene selected for in cattle and AMHs. Upstream of *BRAF*, *SOS1*, selected in domesticated foxes, affects MAPK signaling, bringing about Noonan phenotypes[42]. Noonan-syndrome-like phenotypes are associated with several genes that appear to have undergone selective sweeps in AMH. For instance, *CBL* is located in a region showing signal of a strong selective sweep in AMH compared to Altai Neanderthals[27], and, when mutated, has been shown to give rise to a Noonan syndrome-like disorder[43].

As mentioned above, the AMH-selected neuregulin (NRG) receptor *ERBB4* is part of the ERK/MAPK pathway and negatively regulates ERK via upstream phosphorylation of Raf-1[44]. Loss of function of *Erbb4* in mice has been shown to cause defects in hindbrain cranial neural crest cell pathfinding, including a caudal elongation of the trigeminal and geniculate ganglia[45]. This suggests a plausible role for *ERBB4* in preventing caudal extension in the derived AMH skull. *ERBB4* is one of many neural crest-related genes associated with selective signals in AMH (e.g., *SNAI2*[27], *CITED2*[27], *PRDM10*[46,47]; and other genes discussed in[6]), some of which show fixed or nearly fixed amino acid changes compared to Neanderthals. In addition, *NRG2* was the only gene that was found to be selected in three of the four domesticated species in our study: cat, cattle, and dog. NRG4 was selected in cattle, and NRG3, in AMH. Incidentally, *NRG3* copy number and single nucleotide variants have been associated with Hirschsprung disease[48,49]. This disease is very relevant in the context of domestication, as it affects the neural crest, associated with domestication syndrome[6,13]. Quite a few genes associated with selective sweeps in AMH examined here (among them, *RET, ZEB2*, and *SLIT2*) have been linked to the disease[50,51].

Enhanced ERBB4 signaling has been implicated in Angelman syndrome, an autism spectrum disorder marked by behavioral traits such as increased desire for social interaction, developmental delay, severe speech impairment, and a happy demeanour, although aggressive behavior has sometimes been reported[52–54]. Angelman-syndrome-like phenotypes are frequently associated with genes investigated here. One such observation concerns "the most intriguing variant fixed between dogs and wolves"[29], which is found in the 3'-UTR of *SLC9A6*. This gene encodes sodium/hydrogen exchanger protein 6, which is part of a network related to the plasticity of glutaminergic neurons[55].

Cagan and Blass[29] note that loss-of-function mutations in this gene in humans can lead to Christianson syndrome, also known as "Angelman-like syndrome". Phenotypes typical of these patients include cognitive developmental delays, absence of speech, stereotyped repetitive hand movements, and postnatal microcephaly with a narrow face. Christianson syndrome is frequently characterized by a happy disposition with easily provoked laughter and smiling, an open mouth with excessive drooling and frequent visual fixation on hands. Several of these phenotypes resemble those that distinguish dogs from wolves. It is noteworthy that in pathway analyses carried out on the lists of genes in Tables S1 and S2, as well as the list of genes with amino acid replacement substitutions fixed in AMHs and absent in archaic humans[56], ERKs are among the most significant downstream targets of the interacting selected genes in each. As a final note on the ERK pathway, we highlight the presence of *CACNA1D* in Table S2. *CACNA1D* is a neural cell adhesion molecule that contributes to cell migration via activation of MAPK/ERK signaling[57]. This gene is one of several axon-guidance molecules we identified in our study (see also section 2.3). It is highly expressed in the adrenal glands[58], and, when mutated, gives rise to cerebral palsy/motor disorders[59]. It has been linked to auditory processing[60], and said to be among the positively selected genes in some vocal learners[61].

bioRxiv preprint doi: https://doi.org/10.1101/125799; this version posted April 8, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Glutamate receptors

The glutamate receptor GRIK3 occurs in the AMH, dog, and cattle lists. interacts with other glutamate receptors selected for in the horse (*GRID1*) and cat (*GRIA1/2*). Polymorphisms in *GRIK3* and *GRID1* have been implicated in schizophrenia[62,63], and *GRID1* neighbors *NRG3* (discussed above) at the schizophrenia susceptibility loci 10q22-q23[64]. Developmental delays and craniofacial anomalies associated with a loss of genetic material at the *NRG3* locus accompanied by a gain of material at the *DLGAP1* site have also been reported[65]. *DLGAP1*, a scaffold-protein-coding gene at the postsynaptic density, is in fact selected for in AMHs and in the horse. This gene has been implicated in obsessive-compulsive disorders and interacts significantly with Shank proteins, mutations in which have been linked to autism spectrum disorders with impaired social interaction and communication[66–68]. DLGAP1 interacts with the glutamate receptor GRIK2, implicated in obsessive-compulsive disorders[69].

Previous work[70] already pointed out that genes involved in glutamate metabolism show the greatest population differentiation by whole-genome comparison of dogs and wolves. Although such changes may be implicated in fear response differences between the dog and the wolf populations, Li et al. argue for a role in increasing excitatory synaptic plasticity in dogs rather than reducing fear response. As they point out, changes related to synaptic plasticity may have a significant impact on learning and memory. This is certainly true for cognitive specializations in humans, like language, since glutamate receptors have been shown to be differentially regulated in brain regions associated with vocal learning[71].

Genes selected for in multiple domesticates but not AMH

It is worth considering those genes selected for across domesticates, independently of their selection in AMH, for different reasons. First, our aim here is to explore the extent to which self-domesticating processes in humans may have contributed to our species' anatomical, cognitive, and behavioral make-up. Uncovering strong candidate genes for (albeit often different) domesticating processes in other well-studied domesticated species is a promising way to pursue this goal. The strongest of these candidate genes should be those that have been selected for across different species. Those genes selected for only in certain domesticates, but which strongly interact with genes selected across domesticated species, may prove central to a relevant domesticating process, given that these interactions may shed special light on relevant phenotypic traits. Similarly, certain genes selected for across different domesticates may have trong interactions with other genes that *are* selected for in AMHs. We wish to highlight some of these here (for a full list, see Table S1).

DCC (*DCC Netrin 1 receptor*), an axon-guidance mediator and neural crest-related gene, selected for in both the horse and the cat, interacts strongly with *DSCAM* (*Down Syndrome Cell Adhesion Molecule*), another axon-guidance and neural crest-related gene, selected for in cattle. Of key significance is interaction of *DCC* with the Slit/Robo pathway, especially given the latter's proposed involvement in vocal learning[72,73] and the selection of both *ROBO2* and *SLIT2* in AMHs[47]. (*ROBO1* is also selected for in cattle.) ROBO silences the effect of Netrin 1 to DCC, allowing SLIT2 to bind to this ligand and enabling axon pathfinding in the developing brain[74,75]. *DCC* is involved in the organization of dopaminergic circuits within the cortex[76], and several association studies have identified *DCC* as a promising candidate for schizophrenia[77]. Importantly, an AMH-specific hCONDEL exists in a region upstream of *DCC*, although it is shared with Neanderthals[78]. However, a detailed examination on this gene on both the modern and archaic lines, reveals an accummulation of changes on this gene in AMH. In addition, several genes showing AMH-specific amino-acid substitutions, such as *NOVA1* and *RASA1*, both involved in neuronal development, are known to interact with *DCC*[79,80], and could regulate it in a species-specific fashion. *RASA1* is associated with a strong selective signal in AMH, and has been shown to mediate Netrin 1-induced cortical axon outgrowth and guidance[80]. Together with the glutamate receptor changes discussed above, such modifications may have played an important role in generating aspects of the cognitive profile associated with modern humans, including a full-fledged language-ready brain.

We found several collagen-type genes selected for across domesticates. *COL22A1*, a gene selected for in the horse, significantly interacts with various similar genes selected for in other domesticates, particularly in the cat, including *COL11A1*, selected for in dogs and AMHs. *COL22A1* and *COL11A1* exhibit increased expression in the bone tissue and hippocampus of mice with some of the symptoms of Kleefstra Syndrome (developmental delay, hypotonia, and craniofacial abnormalities), which are often accompanied by autistic symptoms and intellectual disability in humans[81,82].

Archaic-derived alleles

To the best of our knowledge, no comprehensive selective sweep analysis exists for neanderthals. We examined the genes associated with archaic-derived alleles[83] and found no genes in Table S1 display reported archaic-derived alleles. While this could be due to the smaller number of archaic-specific SNCs known at the time of writing, we find this to be an important contrast with the situation that obtains with AMH, in light of the self-domestication hypothesis. It is striking that Castellano et al.[83] highlight genes involved in skeletal development and associated with aggressive phenotypes. We have been able to find data concerning nearly fixed ancestral or derived SNPs in archaic lineages that crop up as variants in modern-day populations. Despite the many confounding factors as to how the relevant mutated genes might interact in different genetic contexts, one might still expect certain archaic-selected SNPs to exhibit somewhat 'underdomesticated' phenotypes when occurring as AMH variants. In this sense, mutations imitating ancestral SNPs found in archaic lineages may be able to tell us a great deal about

bioRxiv preprint doi: https://doi.org/10.1101/125799; this version posted April 8, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

the evolution of our lineage, by allowing us to glimpse some aspects of the ancestral genotype. Among those mutations we found, there is an ancestral S330A mutation of *SLITRK1* that may be involved in obsessive-compulsive disorders like Tourette's Syndrome[84,85]. Different amino acid changes around the site of an AMH-specific derived protein ADSL (A429V) can bring about adenylosuccinate lyase deficiency (R426H; D430N[86]), the symptoms of which include developmental delay, autistic-like traits, aggressiveness, and microcephaly[87].

Discussion

As several scholars have pointed out[5,12,88], there are several routes to domestication. These include the directed pathway (human-assisted amplification of some desired trait in a species), the prey pathway (animals previously hunted in the wild, and subsequently managed in herds), and the commensal pathway (in which animals are drawn to humans by attractive food sources, and over time domesticate themselves). As a result, we should expect genes targeted by a domestication process to differ considerably across species. Nevertheless, reviewing the molecular events associated with domestication reveals common themes, with significant numbers of genes related to brain function and behavior, anatomy, and diet, across domesticates. This is consistent with the view that domestication may be best represented as a spectrum or continuum[89], with a polygenic basis and non-uniform symptomology. This state of affairs is well reflected in the study conducted by Albert et al. (2012)[90], who found significant brain expression changes across domesticates, with the majority of these changes being species-specific.

Because of these findings, we find the overlaps listed in Tables S1-S2 and the associated functions and pathways discussed in the Results section all the more relevant, especially because they converge to a large extent with what is to be expected from the neural-crest-based hypothesis[13] put forth to capture the common mechanistic basis of domestication events. According to Wilkins et al. (2014)[13], a disruption in neural crest developmental programs might be the source of changes spanning multiple organ systems and morphological structures, and the genes examined here seem to broadly support this view. It is quite possible that a neural crest-based explanation won't apply to all domesticates, as suggested in[12], but we find it significant that this hypothesis finds its strongest support in domesticated species like dogs, which have been argued to be self-domesticated[2]. Recall that the goal of the present study was not to provide molecular evidence for a general theory of domestication, but rather to identify domestication-related pathways that could be suggestive of a self-domestication process in AMH. The fact that we find neural-crest-related changed in AMH compared to Neanderthals/Denisovans, and that such changes are also found in another species hypothesized to have undergone a self-domestication process, reinforces our hypothesis that self-domestication took place in our species.

Apart from neural crest-related genes or pathways, we identified common themes pertaining to neuronal development, synaptic plasticity and enhanced learning (categories often mentioned in independent studies on selective sweeps in AMH; e.g.,[47]). These results are in line with claims in other studies on domestication[91–94], where categories like 'neurological process' frequently stand out strongly in gene ontology category enrichment analyses. For scientists interested in the evolution of human cognition, this is of great relevance. It lends credence to claims pairing domestication and a certain type of intelligence[95], and it is not unreasonable to suspect that byproducts of the domestication process such as enhanced sensory-motor perceptual and learning pathways, may provide a foundation for more complex communicative abilities, including vocal learning abilities[7,96]. In a similar vein, we want to point out that among the genes found in Tables S1-2, one finds multiple strong candidates for neurodevelopmental diseases and syndromes (as mentioned for several genes in the Results section). This could be seen as an additional piece of evidence suggestive of a self-domestication process in AMH. A build-up of deleterious alleles is documented across domesticated species when compared to their wild counterparts. Thus, there is a higher frequency of non-synonymous substitutions in the nuclear DNA of domesticated dogs relative to gray wolves[97], and the same is true of their mitochondrial DNA[98]. A higher frequency of non-synonymous substitutions in domesticated yaks compared to the wild yaks has also been reported[99]. This build-up of deleterious alleles has been described as the 'cost of domestication'[100], which, if true, could be a byproduct of self-domestication in AMH, too. For instance, Buffil et al.[101] points out that elevated aerobic metabolism, which may lead to the retention of 'juvenile' neuronal characteristics including elevated synaptic activity, may predispose these neurons to be more sensitive to oxidative stress and thus neurodegenerative diseases.

A study like the present one suffers from several limitations. While we have tried to make our comparisons as fair as possible, we have relied on genomic data that necessarily reflect the current state of the art for the various species we examined. The lists of genes associated with signals of positive selection are derived from the literature, and were generated using different analytical tools. While we have done our best to minimize the number of simplifying assumptions (see Methods), we must point out that even within a single species (e.g., AMH), no two studies completely agree on a definitive list. Indeed, in some cases, they produce lists of very different sizes. In addition, We may have missed important genes of interest due to the lack of information on them in the various databases we consulted. While it is to be hoped that some of these limitations will be overcome in the future, we think that the overlaps discussed in this study should encourage further detailed examination of these genes and the processes they take part in.

We could have been more strict about our notion of convergence, and restrict our attention to genes where the exact

bioRxiv preprint doi: https://doi.org/10.1101/125799; this version posted April 8, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

same difference (e.g., the same amino acid substitution) could be detected between species (for an early attempt along these lines, see[102]). But given that convergent evolution is often hypothesized to occur in the absence of this very strict notion of convergence — for instance, convergent evolution in the domain of vocal learning is related to non-identical changes in *FOXP2* across vocal learners[103] — we feel justified in our approach.

As more data about selective sweeps become available, we think that it will be worth examining the extent to which our results carry over to other species like the bonobo, which has been claimed to be self-domesticated[5]. Likewise, more information about the molecular basis of the famous domestication experiment with the silver fox (*Vulpes vulpes*)[104] should prove valuable. As a matter of fact, the only study focusing on the genetic divergence between foxes that were selected for tame and aggressive behavior[35] reveals an intriguing overlap between a set of SNPs significantly differentiating the two fox strains and among the genes noted in the silver fox data[35] and a number of genes considered to fall within non-coding human accelerated regions (ncHARs)[105]. The most significant among these genes are those implicated in neural tube and forebrain development, and are listed by Racimo (2016)[28] as having been selected from within a single region on chromosome 12 in the AMH Eurasian branch. The region containing all of these genes and no others has been deleted in at least one recorded adult patient with mental retardation and psychiatric problems[106]. Data of this kind have the potential to reinforce the argument presented in this study.

Methods

Data

To identify signatures of a self-domestication process in AMHs, we first constructed a list of genes associated with signs of positive selection in AMH compared to Neanderthals and Denisovans. We then compared this list to the genes independently argued to be associated with positive selection in domesticated species versus their wild counterparts, and examined the overlap between these two gene lists.

For AMH-Neanderthal/Denisovan comparisons, we made use of findings based on high-quality genome reconstructions, specifically: the list of genes in regions of putative selective sweeps, together with pathway and disease annotation in [27, Table S19b.1]; the list of genes from the top 20 candidate regions for the modern human ancestral branch in [28, Table 3]; and the extended list of genomic regions that are predicted to underlie positively selected human specific traits in [47, Supplemental File S1 and S2, Table 2, and Table S7]. We included in our study a range of domesticated species for which detailed genetic information is available. These species offer representative examples of the various routes to domestication[88], as well as different temporal windows for domestication. The species include: dog (*Canis familiaris*)[29,93,107], cat (*Felis catus*)[108], horse (*Equus caballus*)[109], and taurine cattle (*Bos taurus*)[110]. We homogenized the nomenclature across gene sets as best we could.

We also examined other species, including the rabbit (*Oryctolagus cuniculus*)[94], and bonobo (*Pan paniscus*)[111]. In the end, the lists of selected genes for these species (compared to their wild counterparts) were too small to draw any firm conclusion. To help us understand domestication-related changes better, we made use of the comparison of two lines of rats (*Rattus norvegicus*) selected for tame and aggressive behaviour to identify genetic loci that differ between the lines[112], the comparison of gene expression levels in the brains of domesticated and wild animals offered in[90], genomic signatures of domestication in neurogenetic genes in *Drosophila melanogaster*, where neurogenetic genes have been claimed to be associated with signs of positive selection[91], and the genetic divergence between foxes (*Vulpes vulpes*) that were selected for tame and aggressive behavior[35].

For the Great Ape comparison — chimpanzee (*Pan t. troglodytes*), orangutan (*Pongo abelii*), and gorilla (*G. g. gorilla*) — we made use of the data in[113] (positive and balancing selection and selective sweep data in Tables S6, S18(68), S19(69), S20(70), S24(74), and S97).

Methods

In order to test the significance of the overlap between domestication-related genes and genes showing signals of positive selection and selective sweep in AMH, a Hypergeometric Intersection distribution test was performed by using the *R* software[114] and the *R* package `hint` developed in[115]. A Hypergeometric Intersection distribution can be employed to compute the probability of picking an intersection of size v when drawing independently and without replacement from two sets A and B composed of objects of n categories, with a and b number of draws respectively (where $a \neq b$)[115].

As a model of our data we chose as a simplifying assumption $n = 19,500$ as the average number of protein-coding genes for all the species taken into consideration. From the original lists, we thus removed antisense RNA genes (non coding), miRNAs, and other non-coding transcripts/products listed in the original tables.

From this modeled genome, the domesticate pool (comprising cat, dog, cattle, and horse) draws a total of $a = 691$ genes, while the total AMH pool draws $b = 742$ genes. The resulting intersection size (i.e. the number of genes that are positively selected or under selective sweep both in AMH and in one or more domesticate) is $v = 41$. The `hint.test` function was then employed to test the significance of this intersection, obtaining $p < 0.01$.

bioRxiv preprint doi: https://doi.org/10.1101/125799; this version posted April 8, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

A Monte Carlo simulation was performed using Matlab (MathWorks, Natick, MA) to confirm these results. Two random samples, of lengths 691 and 742, were drawn from a pool representing 19,500 genes using Matlab's random number generation function. These simulated draws were performed 1,000,000 times and the percentage of trials in which the intersection was $\geq$41 was calculated. The results revealed that 0.33% of trials had intersections of this size.

Intersections between domesticates were tested, revealing a significant overlap between genes selected in dog and in cattle ($a = 229$, $b = 78$, $v = 5$, $p < 0.01$). Furthermore, tests between AMH and each domesticate showed significant overlaps with dog ($v = 15$, $p < 0.05$) and with cattle ($v = 9$, $p < 0.01$).

Since we pooled data for positive selection and selective sweep in AMH from different sources, intersection tests were carried out between the domestication pool and the pool of each AMH dataset used in this study. A significant intersection was found with the data in[27] ($a = 691$, $b = 108$, $v = 9$; $p < 0.05$) and with the combined data from[27] and[28] ($a = 691$, $b = 419$, $v = 24$, $p < 0.05$).

Overlaps with domesticates were checked for Great Apes, using data from[113]. For chimpanzee (*Pan t. troglodytes*), $b = 415$ with $v = 16$; for orangutan (*Pongo abelii*), $b = 500$ with $v = 20$; for gorilla (*G. g. gorilla*), $b = 426$ with $v = 12$. The tests yielded non-significant results for all these intersections. Monte Carlo simulations, performed as described above, *mutatis mutandis*, showed that intersections of these sizes occurred in a large fraction of trials (40.11% of trials for chimpanzee; 32% for orangutan; 82.89% for gorilla). As in the case of AMH, overlaps with individual domesticates were tested, with no significant results.

We then examined the functions of the genes in Tables S1-3, paying close attention to the pathways in which they are involved, and to their interactions with other genes already highlighted in the domestication literature. In addition to performing an exhaustive PubMed (http://www.ncbi.nlm.nih.gov/pubmed) search on each of the genes, we drew on the information available in Genecards (http://genecards.org), Uniprot (http://www.uniprot.org/), String 10.0 (http://string-db.org), and Biogrid 3.4 (http://thebiogrid.org) to identify potential protein-protein interactions and Gene Ontology category enrichment signals. Additionally, we studied the networks and functional analyses generated by feeding our gene lists into QIAGEN's Ingenuity Pathway Analysis software (https://www.qiagen.com/ingenuity).

Furthermore, we gathered information about the expression patterns of these genes, concentrating on those genes with relatively high expression in tissues such as brain, bone, and adrenal glands. For this, we relied on the following resources: Brainspan (http://www.brainspan.org), Human Brain Transcriptome (http://hbatlas.org), Bgee (http://bgee.org), Proteomics DB (https://proteomicsdb.org), Human Protein Atlas (http://www.proteinatlas.org), Gene Enrichment Profiler (http://xavierlab2.mgh.harvard.edu/EnrichmentProfiler/index.html), and GTex (http://www.gtexportal.org). For the information presented in the Supplementary Material, we consulted the following databases: KEGG Pathways and Disease (http://www.kegg.jp/kegg/), PANTHER (http://www.pantherdb.org), Reactome Pathway Database (http://www.reactome.org), OMIM (http://omim.org), and MalaCards (http://www.malacards.org/).

References

1. Thomas, J. *Self-domestication and language evolution*. Ph.D. thesis, University of Edinburgh (2014).
2. Hare, B. Survival of the friendliest: Homo sapiens evolved via selection for prosociality. *Annu. Rev. Psychol.* **68** (2016).
3. Tomasello, M. *The cultural origins of human cognition* (Harvard university press, 2009).
4. Hare, B. & Tomasello, M. Human-like social skills in dogs? *Trends cognitive sciences* **9**, 439–444 (2005).
5. Hare, B., Wobber, V. & Wrangham, R. The self-domestication hypothesis: evolution of bonobo psychology is due to selection against aggression. *Animal Behav.* **83**, 573–585 (2012).
6. Benítez-Burraco, A., Theofanopoulou, C. & Boeckx, C. Globularization and domestication. *Topoi* 1–14 (2016).
7. Boeckx, C. Evolution of language. In Kaas, J. (ed.) *Evolution of Nervous Systems, 2nd ed., vol. 4*, 325–339 (Elsevier, London, 2017).
8. Okanoya, K. Sexual communication and domestication may give rise to the signal complexity necessary for the emergence of language: An indication from songbird studies. *Psychon. Bull. & Rev.* 1–5 (2017).
9. Kirby, S. Culture and biology in the origins of linguistic structure. *Psychon. Bull. & Rev.* (2017). Springer Compact Gold OA.
10. Cieri, R. L., Churchill, S. E., Franciscus, R. G., Tan, J. & Hare, B. Craniofacial feminization, social tolerance, and the origins of behavioral modernity. *Curr. Anthropol.* **55**, 419–443 (2014).
11. Stringer, C. The origin and evolution of homo sapiens. *Phil. Trans. R. Soc. B* **371**, 20150237 (2016).
12. Sánchez-Villagra, M. R., Geiger, M. & Schneider, R. A. The taming of the neural crest: a developmental perspective on the origins of morphological covariation in domesticated mammals. *Open Sci.* **3**, 160107 (2016).

bioRxiv preprint doi: https://doi.org/10.1101/125799; this version posted April 8, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

13. Wilkins, A. S., Wrangham, R. W. & Fitch, W. T. The "domestication syndrome" in mammals: a unified explanation based on neural crest cell behavior and genetics. *Genet.* **197**, 795–808 (2014).

14. Hublin, J.-J., Neubauer, S. & Gunz, P. Brain ontogeny and life history in pleistocene hominins. *Phil. Trans. R. Soc. B* **370**, 20140062 (2015).

15. Lacruz, R. *et al.* Facial ontogeny in neanderthals and their ancestors. *Nat. Commun.* (2015).

16. Maureille, B. & Bar, D. The premaxilla in neandertal and early modern children: ontogeny and morphology. *J. human evolution* **37**, 137–152 (1999).

17. de León, M. S. P. *et al.* Neanderthal brain size at birth provides insights into the evolution of human life history. *Proc. Natl. Acad. Sci.* **105**, 13764–13768 (2008).

18. Zilberman, U. & Smith, P. A comparison of tooth structure in neanderthals and early homo sapiens sapiens: a radiographic study. *J. anatomy* **180**, 387 (1992).

19. Carter, K. & Worthington, S. Morphologic and demographic predictors of third molar agenesis a systematic review and meta-analysis. *J. dental research* **94**, 886–894 (2015).

20. Trut, L., Oskina, I. & Kharlamova, A. Animal evolution during domestication: the domesticated fox as a model. *Bioessays* **31**, 349–360 (2009).

21. Künzl, C. & Sachser, N. The behavioral endocrinology of domestication: a comparison between the domestic guinea pig (cavia apereaf. porcellus) and its wild ancestor, the cavy (cavia aperea). *Horm. Behav.* **35**, 28–37 (1999).

22. Schaefer, K., Fink, B., Mitteroecker, P., Neave, N. & Bookstein, F. L. Visualizing facial shape regression upon 2nd to 4th digit ratio and testosterone. *Coll. antropologicum* **29**, 415–419 (2005).

23. Nelson, E., Rolian, C., Cashmore, L. & Shultz, S. Digit ratios predict polygyny in early apes, ardipithecus, neanderthals and early modern humans but not in australopithecus. *Proc. Royal Soc. Lond. B: Biol. Sci.* **278**, 1556–1563 (2011).

24. Knight, C., Power, C. & Watts, I. The human symbolic revolution: a darwinian account. *Camb. archaeological journal* **5**, 75–114 (1995).

25. Orlando, L. Ancient dna and the study of animal domestication. *Annu. Rev. Animal Biosci.* **5** (2016).

26. Llamas, B., Willerslev, E. & Orlando, L. Human evolution: a tale from ancient genomes. *Phil. Trans. R. Soc. B* **372**, 20150484 (2017).

27. Prüfer, K. *et al.* The complete genome sequence of a neanderthal from the altai mountains. *Nat.* **505**, 43–49 (2014).

28. Racimo, F. Testing for ancient selection using cross-population allele frequency differentiation. *Genet.* **202**, 733–750 (2016).

29. Cagan, A. & Blass, T. Identification of genomic variants putatively targeted by selection during dog domestication. *BMC evolutionary biology* **16**, 1 (2016).

30. Argente, J. *et al.* Defective minor spliceosome mrna processing results in isolated familial growth hormone deficiency. *EMBO molecular medicine* e201303573 (2014).

31. Gucev, Z. *et al.* Severe isolated growth hormone deficiency and myopathy in two brothers with rnpc3 mutation. *ESPE Abstr.* (2015).

32. Rada-Iglesias, A. *et al.* Epigenomic annotation of enhancers predicts transcriptional regulators of human neural crest. *Cell stem cell* **11**, 633–648 (2012).

33. Simões-Costa, M. & Bronner, M. E. Establishing neural crest identity: a gene regulatory recipe. *Dev.* **142**, 242–257 (2015).

34. Rosa, A. & Brivanlou, A. H. A regulatory circuitry comprised of mir-302 and the transcription factors oct4 and nr2f2 regulates human embryonic stem cell differentiation. *The EMBO journal* **30**, 237–248 (2011).

35. Johnson, J. L. *et al.* Genotyping-by-sequencing (gbs) detects genetic structure and confirms behavioral qtl in tame and aggressive foxes (vulpes vulpes). *PloS one* **10**, e0127013 (2015).

36. Gau, S. S.-F., Liao, H.-M., Hong, C.-C., Chien, W.-H. & Chen, C.-H. Identification of two inherited copy number variants in a male with autism supports two-hit and compound heterozygosity models of autism. *Am. J. Med. Genet. Part B: Neuropsychiatr. Genet.* **159**, 710–717 (2012).

37. Sweatt, J. D. The neuronal map kinase cascade: a biochemical signal integration system subserving synaptic plasticity and memory. *J. neurochemistry* **76**, 1–10 (2001).

bioRxiv preprint doi: https://doi.org/10.1101/125799; this version posted April 8, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

38. Cesarini, L. *et al.* Cognitive profile of disorders associated with dysregulation of the ras/mapk signaling cascade. *Am. J. Med. Genet. Part A* **149**, 140–146 (2009).

39. Newbern, J. *et al.* Mouse and human phenotypes indicate a critical conserved role for erk2 signaling in neural crest development. *Proc. Natl. Acad. Sci.* **105**, 17115–17120 (2008).

40. Keyte, A. & Hutson, M. R. The neural crest in cardiac congenital anomalies. *Differ.* **84**, 25–40 (2012).

41. Sarkozy, A. *et al.* Germline braf mutations in noonan, leopard, and cardiofaciocutaneous syndromes: molecular diversity and associated phenotypic spectrum. *Hum. mutation* **30**, 695–702 (2009).

42. Roberts, A. E. *et al.* Germline gain-of-function mutations in sos1 cause noonan syndrome. *Nat. genetics* **39**, 70–74 (2007).

43. Martinelli, S. *et al.* Heterozygous germline mutations in the cbl tumor-suppressor gene cause a noonan syndrome-like phenotype. *The Am. J. Hum. Genet.* **87**, 250–257 (2010).

44. Hatakeyama, M. *et al.* A computational model on the modulation of mitogen-activated protein kinase (mapk) and akt pathways in heregulin-induced erbb signalling. *Biochem. J.* **373**, 451–463 (2003).

45. Golding, J. P., Trainor, P., Krumlauf, R. & Gassmann, M. Defects in pathfinding by cranial neural crest cells in mice lacking the neuregulin receptor erbb4. *Nat. cell biology* **2**, 103–109 (2000).

46. Racimo, F., Kuhlwilm, M. & Slatkin, M. A test for ancient selective sweeps and an application to candidate sites in modern humans. *Mol. biology evolution* **31**, 3344–3358 (2014).

47. Peyrégne, S., Dannemann, M. & Prüfer, K. Detecting ancient positive selection in humans using extended lineage sorting. *bioRxiv* 092999 (2016).

48. Tang, C. S.-M. *et al.* Genome-wide copy number analysis uncovers a new HSCR gene: *NRG3*. *PLoS Genet.* **8**, 10.1371 (2012).

49. Yang, J. *et al.* Exome sequencing identified *NRG3* as a novel susceptible gene of Hirschsprung's disease in a Chinese population. *Mol Neurobiol* **47**, 957–966 (2013).

50. Jiang, Q., Ho, Y.-Y., Hao, L., Berrios, C. N. & Chakravarti, A. Copy number variants in candidate genes are genetic modifiers of hirschsprung disease. *PLoS One* **6**, e21219 (2011).

51. Tang, W. *et al.* Slit2/robo1-mir-218-1-ret/plag1: a new disease pathway involved in hirschsprung's disease. *J. cellular molecular medicine* **19**, 1197–1207 (2015).

52. Clayton-Smith, J. & Laan, L. Angelman syndrome: a review of the clinical and genetic aspects. *J. Med. Genet.* **40**, 87–95 (2003).

53. Bird, L. M. Angelman syndrome: review of clinical and molecular aspects. *Appl. Clin. Genet.* **7** (2014).

54. Summers, J. A., Allison, D., Lynch, P. & Sandier, L. Behaviour problems in angelman syndrome. *J. Intellect. Disabil. Res.* **39**, 97–106 (1995).

55. Gilfillan, G. D. *et al.* Slc9a6 mutations cause x-linked mental retardation, microcephaly, epilepsy, and ataxia, a phenotype mimicking angelman syndrome. *The Am. J. Hum. Genet.* **82**, 1003–1010 (2008).

56. Pääbo, S. The human condition—a molecular approach. *Cell* **157**, 216–226 (2014).

57. Shi, Y. *et al.* Neural cell adhesion molecule modulates mesenchymal stromal cell migration via activation of mapk/erk signaling. *Exp. cell research* **318**, 2257–2267 (2012).

58. Azizan, E. A. *et al.* Somatic mutations in atp1a1 and cacna1d underlie a common subtype of adrenal hypertension. *Nat. genetics* **45**, 1055–1060 (2013).

59. Scholl, U. I. *et al.* Somatic and germline cacna1d calcium channel mutations in aldosterone-producing adenomas and primary aldosteronism. *Nat. genetics* **45**, 1050–1054 (2013).

60. Satheesh, S. V. *et al.* Retrocochlear function of the peripheral deafness gene cacna1d. *Hum. molecular genetics* dds217 (2012).

61. Nam, K. *et al.* Molecular evolution of genes in avian genomes. *Genome biology* **11**, R68 (2010).

62. Begni, S. *et al.* Association between the ionotropic glutamate receptor kainate 3 (grik3) ser310ala polymorphism and schizophrenia. *Mol. psychiatry* **7**, 416–418 (2002).

63. Guo, S.-Z. *et al.* A case-control association study between the grid1 gene and schizophrenia in the chinese northern han population. *Schizophr. research* **93**, 385–390 (2007).

bioRxiv preprint doi: https://doi.org/10.1101/125799; this version posted April 8, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

64. Chen, P.-L. *et al.* Fine mapping on chromosome 10q22-q23 implicates neuregulin 3 in schizophrenia. *The Am. J. Hum. Genet.* **84**, 21–34 (2009).

65. Sahoo, T. *et al.* Copy number variants of schizophrenia susceptibility loci are associated with a spectrum of speech and developmental delays and behavior problems. *Genet. Medicine* **13**, 868–880 (2011).

66. Bozdagi, O. *et al.* Haploinsufficiency of the autism-associated shank3 gene leads to deficits in synaptic function, social interaction, and social communication. *Mol. autism* **1**, 1 (2010).

67. Stewart, S. E. *et al.* Genome-wide association study of obsessive-compulsive disorder. *Mol. psychiatry* **18**, 788–798 (2013).

68. Won, H. *et al.* Autistic-like social behaviour in shank2-mutant mice improved by restoring nmda receptor function. *Nat.* **486**, 261–265 (2012).

69. Mattheisen, M. *et al.* Genome-wide association study in obsessive-compulsive disorder: results from the ocgas. *Mol. psychiatry* **20**, 337–344 (2015).

70. Li, Y. *et al.* Domestication of the dog from the wolf was promoted by enhanced excitatory synaptic plasticity: a hypothesis. *Genome biology evolution* **6**, 3115–3121 (2014).

71. Jarvis, E. D. Evolution of brain pathways for vocal learning in birds and humans. *Birdsong, speech, language* 63–108 (2014).

72. Wang, R. *et al.* Convergent differential regulation of slit-robo axon guidance genes in the brains of vocal learners. *J. Comp. Neurol.* **523**, 892–906 (2015).

73. Mozzi, A. *et al.* The evolutionary history of genes involved in spoken and written language: beyond foxp2. *Sci. reports* **6** (2016).

74. Stein, E. & Tessier-Lavigne, M. Hierarchical organization of guidance receptors: silencing of netrin attraction by slit through a robo/dcc receptor complex. *Sci.* **291**, 1928–1938 (2001).

75. Dickinson, R. E. & Duncan, W. C. The slit–robo pathway: a regulator of cell function with implications for the reproductive system. *Reproduction* **139**, 697–704 (2010).

76. Manitt, C. *et al.* The netrin receptor dcc is required in the pubertal organization of mesocortical dopamine circuitry. *The J. Neurosci.* **31**, 8381–8394 (2011).

77. Grant, A., Fathalli, F., Rouleau, G., Joober, R. & Flores, C. Association between schizophrenia and genetic variation in dcc: a case–control study. *Schizophr. research* **137**, 26–31 (2012).

78. McLean, C. Y. *et al.* Human-specific loss of regulatory dna and the evolution of human-specific traits. *Nat.* **471**, 216–219 (2011).

79. Leggere, J. C. *et al.* Nova regulate dcc alternative splicing during neuronal migration and axon guidance in the spinal cord. *eLife* **5**, e14264 (2016).

80. Antoine-Bertrand, J. *et al.* p120rasgap protein mediates netrin-1 protein-induced cortical axon outgrowth and guidance. *J. Biol. Chem.* **291**, 4589–4602 (2016).

81. Balemans, M. C. *et al.* Reduced euchromatin histone methyltransferase 1 causes developmental delay, hypotonia, and cranial abnormalities associated with increased bone gene expression in kleefstra syndrome mice. *Dev. biology* **386**, 395–407 (2014).

82. Willemsen, M. H. *et al.* Update on kleefstra syndrome. *Mol. syndromology* **2**, 202–212 (2012).

83. Castellano, S. *et al.* Patterns of coding variation in the complete exomes of three neandertals. *Proc. Natl. Acad. Sci.* **111**, 6666–6671 (2014).

84. Ozomaro, U. *et al.* Characterization of slitrk1 variation in obsessive-compulsive disorder. *PloS one* **8**, e70376 (2013).

85. Alexander, J. *et al.* Targeted re-sequencing approach of candidate genes implicates rare potentially functional variants in tourette syndrome etiology. *Front. Neurosci.* **10** (2016).

86. Kmoch, S. *et al.* Human adenylosuccinate lyase (adsl), cloning and characterization of full-length cdna and its isoform, gene structure and molecular basis for adsl deficiency in six patients. *Hum. molecular genetics* **9**, 1501–1513 (2000).

87. Jurecka, A., Zikanova, M., Kmoch, S. & Tylki-Szymańska, A. Adenylosuccinate lyase deficiency. *J. inherited metabolic disease* **38**, 231–242 (2015).

88. Zeder, M. A. Pathways to animal domestication. *Biodivers. agriculture: Domest. evolution, sustainability* 227 (2012).

bioRxiv preprint doi: https://doi.org/10.1101/125799; this version posted April 8, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

89. Vigne, J.-D. The origins of animal domestication and husbandry: a major change in the history of humanity and the biosphere. *Comptes rendus biologies* **334**, 171–181 (2011).

90. Albert, F. W. *et al.* A comparison of brain gene expression levels in domesticated and wild animals. *PLoS Genet.* **8**, e1002962 (2012).

91. Stanley, C. E. & Kulathinal, R. J. Genomic signatures of domestication on neurogenetic genes in drosophila melanogaster. *BMC evolutionary biology* **16**, 1 (2016).

92. Wang, G.-d. *et al.* The genomics of selection in dogs and the parallel evolution between dogs and humans. *Nat. communications* **4**, 1860 (2013).

93. Freedman, A. H. *et al.* Demographically-based evaluation of genomic regions under selection in domestic dogs. *PLoS Genet.* **12**, e1005851 (2016).

94. Carneiro, M. *et al.* Rabbit genome analysis reveals a polygenic basis for phenotypic change during domestication. *Sci.* **345**, 1074–1079 (2014).

95. Hare, B. & Woods, V. *The genius of dogs: how dogs are smarter than you think* (Penguin, 2013).

96. Okanoya, K. Behavioural factors governing song complexity in bengalese finches. *Int. J. Comp. Psychol.* **25** (2012).

97. Cruz, F., Vilà, C. & Webster, M. T. The legacy of domestication: accumulation of deleterious mutations in the dog genome. *Mol. biology evolution* **25**, 2331–2336 (2008).

98. Björnerfeldt, S., Webster, M. T. & Vilà, C. Relaxation of selective constraint on dog mitochondrial dna following domestication. *Genome Res.* **16**, 990–994 (2006).

99. Wang, Z. *et al.* Domestication relaxed selective constraints on the yak mitochondrial genome. *Mol. biology evolution* **28**, 1553–1556 (2011).

100. Lu, J. *et al.* The accumulation of deleterious mutations in rice genomes: a hypothesis on the cost of domestication. *Trends Genet.* **22**, 126–131 (2006).

101. Bufill, E., Agustí, J. & Blesa, R. Human neoteny revisited: the case of synaptic plasticity. *Am. J. Hum. Biol.* **23**, 729–739 (2011).

102. Wang, R. *Dissecting the Genetic Basis of Convergent Complex Traits Based on Molecular Homoplasy*. Ph.D. thesis, Duke University (2011).

103. Webb, D. M. & Zhang, J. Foxp2 in song-learning birds and vocal-learning mammals. *J. Hered.* **96**, 212–216 (2005).

104. Trut, L. Early canid domestication: The farm-fox experiment foxes bred for tamability in a 40-year experiment exhibit remarkable transformations that suggest an interplay between behavioral genetics and development. *Am. Sci.* **87**, 160–169 (1999).

105. Capra, J. A., Erwin, G. D., McKinsey, G., Rubenstein, J. L. & Pollard, K. S. Many human accelerated regions are developmental enhancers. *Phil. Trans. R. Soc. B* **368**, 20130025 (2013).

106. Gijsbers, A. C. *et al.* Additional cryptic cnvs in mentally retarded patients with apparently balanced karyotypes. *Eur. journal medical genetics* **53**, 227–233 (2010).

107. Axelsson, E. *et al.* The genomic signature of dog domestication reveals adaptation to a starch-rich diet. *Nat.* **495**, 360–364 (2013).

108. Montague, M. J. *et al.* Comparative analysis of the domestic cat genome reveals genetic signatures underlying feline biology and domestication. *Proc. Natl. Acad. Sci.* **111**, 17230–17235 (2014).

109. Schubert, M. *et al.* Prehistoric genomes reveal the genetic foundation and cost of horse domestication. *Proc. Natl. Acad. Sci.* **111**, E5661–E5669 (2014).

110. Qanbari, S. *et al.* Classic selective sweeps revealed by massive sequencing in cattle. *PLoS Genet.* **10**, e1004148 (2014).

111. Prüfer, K. *et al.* The bonobo genome compared with the chimpanzee and human genomes. *Nat.* **486**, 527–531 (2012).

112. Heyne, H. O. *et al.* Genetic influences on brain gene expression in rats selected for tameness and aggression. *Genet.* **198**, 1277–1290 (2014).

113. Cagan, A. *et al.* Natural selection in the great apes. *Mol. Biol. Evol.* **33**, 3268–3283 (2016).

114. R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria (2013). URL `http://www.R-project.org/`. ISBN 3-900051-07-0.

115. Kalinka, A. T. The probability of drawing intersections: extending the hypergeometric distribution. *arXiv preprint arXiv:1305.0717* (2013).

bioRxiv preprint doi: https://doi.org/10.1101/125799; this version posted April 8, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Acknowledgements

We acknowledge the help of Nina Riddell in generating Figure 2. The drawing for Figure 1 was done by Francisco Peña. We thank Kay Sušelj for assistance with the Monte Carlo simulations.

Funding statement

CB acknowledges the financial support from the Spanish Ministry of Economy and Competitiveness (grant FFI2016-78034-C2-1-P), a Marie Curie International Reintegration Grant from the European Union (PIRG-GA-2009-256413), research funds from the Fundació Bosch i Gimpera, and from the Generalitat de Catalunya (2014-SGR-200). CTh and TOR acknowledge support from the Generalitat de Catalunya in the form of doctoral (FI) fellowships.

Author contributions statement

CB ad CTh formulated the initial hypothesis. SG and CB designed the present study. SG, FD, CB, PTM, and BS selected the methods and performed the statistical analyses. Data collection: SG focused on genomic information related to the dog; PTM, to the cat; TOR, to cattle; AM, to the horse. CTh collected data on the bonobo and SA collected data on the rabbit, which were eventually not included in the present report. CB and CTh produced a first draft of Sections 1 and 3; CB, TOR and SG produced a first draft of Section 2. CB, SG, and AM produced a first draft of Section 4. CB and BS thoroughly revised the drafts and produced the final version of the manuscript, with input from all authors. AM, SG, TOR, BS, and CB prepared the supplementary material. CTh and SA prepared the illustrations. CB coordinated the study.

Additional information

The authors declare no competing financial interests.

bioRxiv preprint doi: https://doi.org/10.1101/125799; this version posted April 8, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

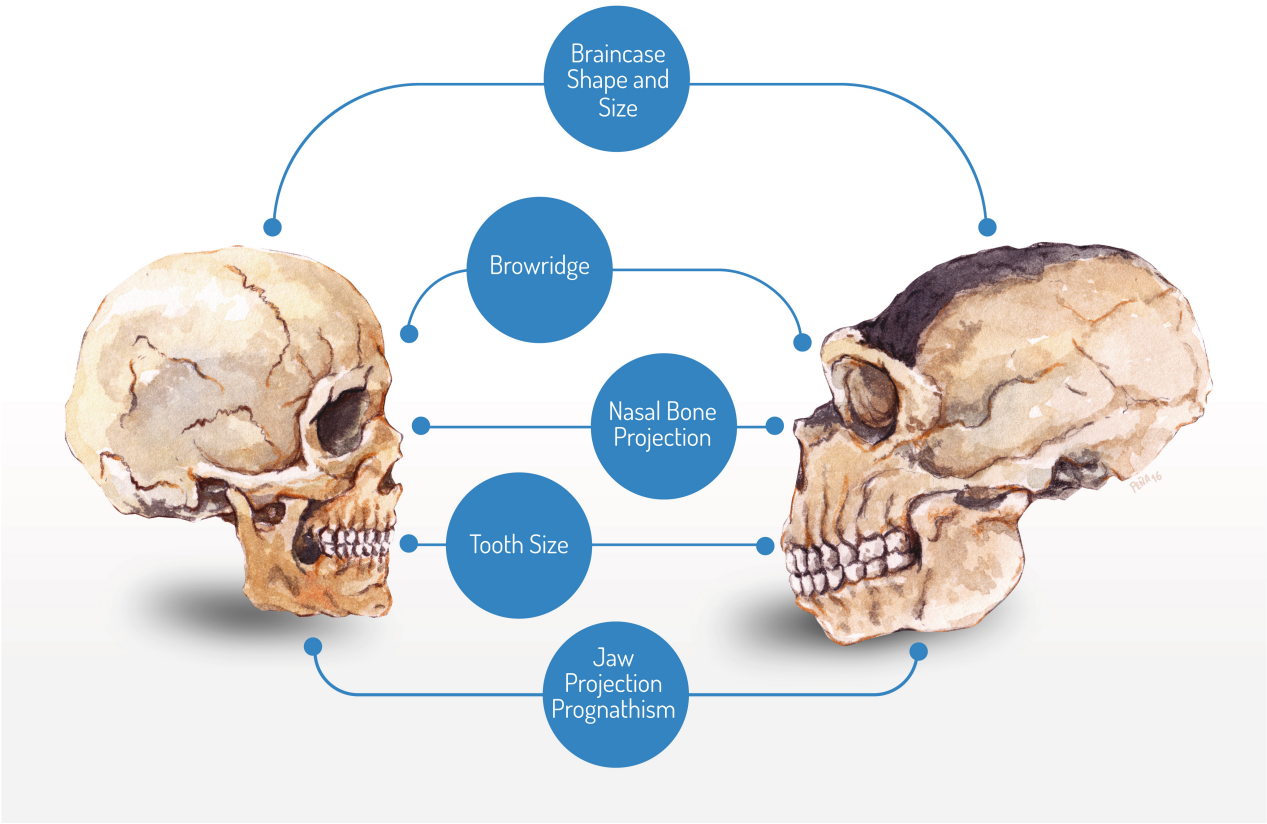

Figure 1. Salient craniofacial differences between Neanderthal and AMH, suggestive of domestication in the latter

bioRxiv preprint doi: https://doi.org/10.1101/125799; this version posted April 8, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

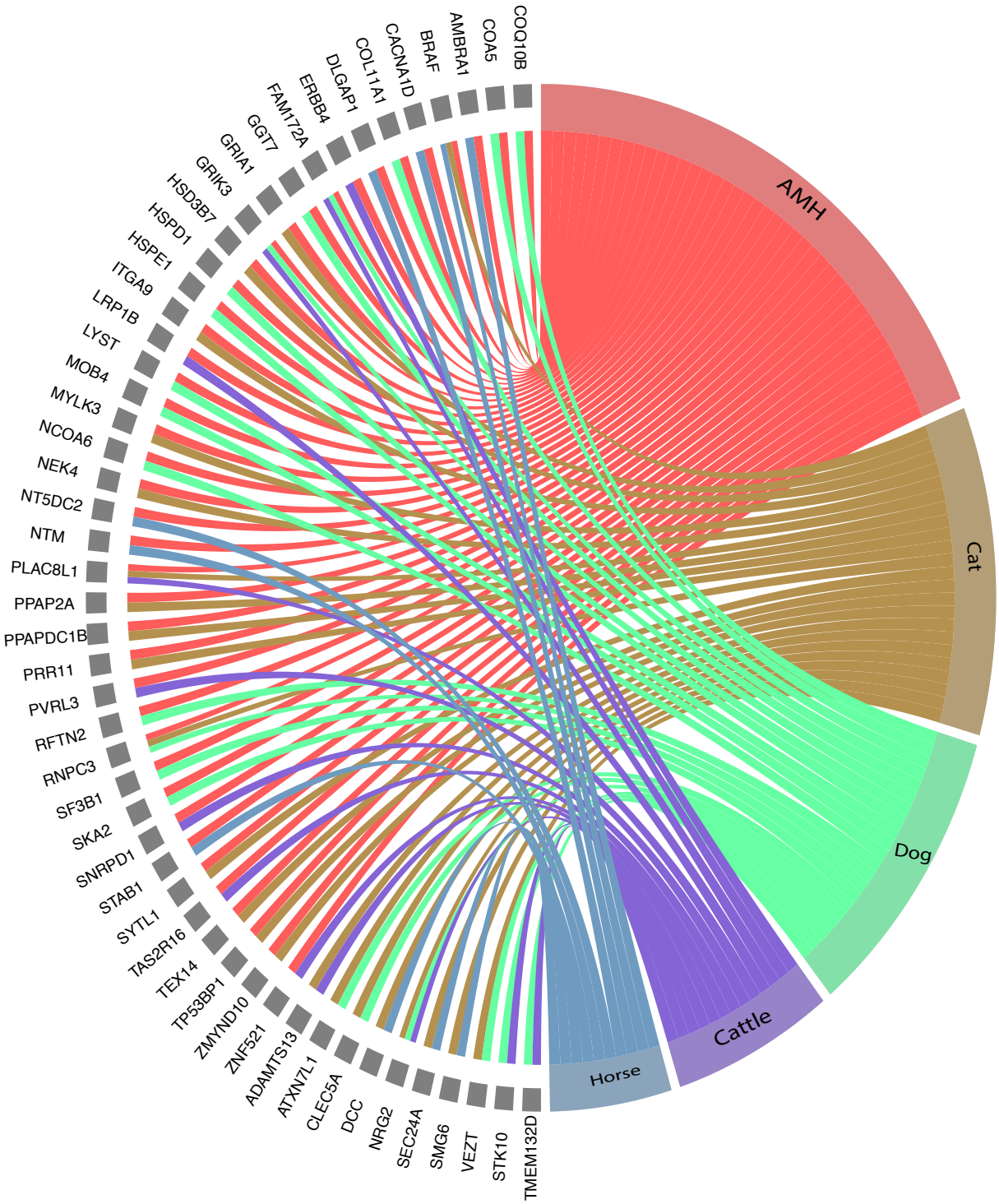

Figure 2. Graphical representation of the overlap identified among gene lists discussed in the text