Abstract

Angiotensin converting enzyme (ACE) is essential for control of blood pressure. The human ACE gene contains an intronic Alu indel (I/D) polymorphism that has been associated with variation in serum enzyme levels, although the functional mechanism has not been identified. The polymorphism has also been associated with cardiovascular disease, type II diabetes, renal disease and elite athleticism. We have characterized the ACE gene in horses of breeds selected for differing physical abilities. The equine gene has a similar structure to that of all known mammalian ACE genes. Nine common single nucleotide polymorphisms (SNPs) discovered in pooled DNA were found to be inherited in nine haplotypes. Three of these SNPs were located in intron 16, homologous to that containing the Alu polymorphism in the human. A highly conserved 18 bp sequence, also within that intron, was identified as being a potential binding site for the transcription factors Oct-1, HFH-1 and HNF-3β, and lies within a larger area of higher than normal homology. This putative regulatory element may contribute to regulation of the documented inter-individual variation in human circulating enzyme levels, for which a functional mechanism is yet to be defined. Two equine SNPs occurred within the conserved area in intron 16, although neither of them disrupted the putative binding site. We propose a possible regulatory mechanism of the ACE gene in mammalian species which was previously unknown. This advance will allow further analysis leading to a better understanding of the mechanisms underpinning the associations seen between the human Alu polymorphism and enzyme levels, cardiovascular disease states and elite athleticism.

Introduction

Angiotensin converting enzyme (ACE) is an essential component of the renin-angiotensin system and plays an important role in the control of blood pressure, renal function and male fertility [1]. The presence of a 287 bp Alu insertion/deletion (I/D) polymorphism with a high minor allele frequency (0.4–0.47) within the ACE gene [2,3], combined with the intrinsic function of the enzyme, has resulted in over 500 association studies between the human ACE gene and a wide range of disorders, most notably cardiovascular disease, type II diabetes and related renal disease [4–7]. More than 20 studies have explored a possible association with extreme athleticism, with conflicting results. In many studies, the insertion (I) allele has been associated with anabolic response to training and elite endurance performance, whilst the deletion (D) allele was associated with sprint or power performance [8–12], and both variants have been associated with response to strength training [13,14]. However, some studies have found no connection with athletic performance [15–18].

The I/D polymorphism, which is found in intron 16, appears to account for 28–47% of the inter-individual variation in serum ACE levels, either due to increased expression of the D allele mRNA, or lower stability of the I allele mRNA [3,19,20]. Circulating enzyme levels were found to be influenced by the action of a major codominant gene, with adults homozygous for the I allele having significantly lower circulating enzyme levels than homozygotes for the D allele. Heterozygotes fall in between these levels [3,19]. The suggestion that an unidentified intronic silencer element is eliminated by the deletion variant, and thus increases observed circulating enzyme levels, has been ruled out [21]. Other studies have hypothesised that other polymorphisms are responsible for the effects attributed to the I/D, and instead indicate that potentially two functional variants exist, probably in the 3’ region of the gene, accounting for these effects [22–24]. However, these studies were unable to identify the functional variant(s) and only investigated a selection of known ACE polymorphisms. Furthermore, the mode of action of alternate allelic forms of the ACE gene on variation in circulating enzyme levels in addition to performance is yet to be elucidated.

The human ACE gene spans 21308 bp [GenBank:NG_011648] of which 4422 bp comprises coding sequence across 26 exons [25]. It encodes two commonly expressed isozymes, the larger of which is membrane bound and primarily found in both epithelial and endothelial cells. In particular, vascular cells from the brain and...
lung produce large amounts of ACE, as do the brush border cells of kidney tubules, while all mammalian endothelial cells appear to produce ACE (endothelial or somatic ACE, sACE) [26]. Vascular endothelial cells also release a circulating form of the enzyme by cleaving it from the membrane bound tail [27,28]. This large isozyme is transcribed from exons 1 to 26, excluding exon 13. The smaller ACE variant is only expressed after puberty in the germinal cells of the testes, is encoded by exons 13 to 26 through initiation of a separate promoter in intron 12, and is known as testicular ACE (tACE). A feature of the ACE gene is a high degree of homology between two distinct regions of the gene, namely exons 4 to 11 (region 1) and 17 to 24 (region 2). The exons in regions 1 and 2 have conserved codon phases, are up to 80% similar, are identical lengths, and are likely the result of a gene duplication event that occurred before mammalian radiation [29].

The horse presents itself a suitable species to study the ACE gene in a biological model of athletic performance. Horses are unique in livestock breeding in that they have almost exclusively been selected for athletic performance, ranging from extreme speed, endurance, or heavy draught performance. Additionally, horses have been shown to exhibit similar interanimal variation in circulating ACE levels as observed in humans [30]. We present here a comprehensive characterization of the equine angiotensin converting enzyme gene in comparison with other mammalian ACE genes, in particular the human, to shed light on the possible mode of action of ACE gene polymorphisms on circulating enzyme levels and thus performance in an athletic animal model.

Materials and Methods

Ethics Statement

All horses sampled for this project were covered by Animal Ethics protocols N02/5-99/1/2946 and N00/3-2002/1/3353, as approved by the University of Sydney Animal Ethics Committee.

Animals

Horses of seven different breeds including racing Thoroughbreds (TB), Standardbreds (SB), Draught/Heavy Horses (HH) (Shires and Clydesdales), endurance Arabians (AR), Quarter Horses (QH) and ponies (PO) were included in this study. The TBs were selected based on minimal degree of relatedness according to pedigree; the ARs were selected for successful endurance performance at distances over 90 km; and for all other breed samples animals were selected at random. Blood samples were collected by venipuncture of the jugular vein for extraction of DNA.

DNA and RNA Extraction

DNA was extracted using a QIAGEN Plasmid Midi Kit (Qiagen, Hilden, Germany) from BAC clone 801F9, which was identified as containing the equine ACE gene. This clone was supplied from the INRA horse BAC library by Dr. Francois Piumi, INRA [31]. Equine DNA was extracted from fresh blood as previously described [32] or as per manufacturer’s instructions from frozen blood samples using a QIAamp® DNA Blood Mini Kit (Qiagen). RNA was extracted from blood using an RNeasy® Mini Kit (Qiagen).

The DNA samples were combined to create 3 pools, composed of TB (n = 10), AR (n = 14), and mixed breeds (MB) (including 2 each of HH, SB, TB, QH and PO) for the discovery of polymorphisms. Common polymorphisms discovered in the pools were also typed across a panel of 40 horses (10 each of TB, AR, SB and HH) to obtain allele frequency and haplotype information.

Primers

Since the equine ACE coding sequence was not available at the commencement of this study, primers were selected based on the aligned cDNA sequences of the human, rabbit, rat and chicken genes [GenBank:J04144, X62551, AF201332 and L40175]. One of three standard M13 tail sequences was synthesized on the 5’ end of the primers (as listed in Table S1), to allow incorporation of a fluorescent label into the PCR product [33].

PCR and Sequencing

Reverse transcriptase reaction PCR was carried out with 2 μg total RNA and reagents from Promega (Madison, WI, USA) and Life Technologies (Grand Island, NY, USA). The cDNA was diluted to a 1 in 5 dilution and 1 μL used as a template in PCR with a total volume of 25 μL.

PCR reactions for genomic, BAC and cDNA were performed in 25 μL volumes containing 20 ng of purified DNA and reagents from Fisher Biotech (Perth, Australia). Primer sequences for determining the gene sequence and screening the DNA pools are shown in Table S1.

PCR product was cleaned up using ExoSapIT (GE Healthcare Life Sciences, Buckinghamshire, England) or JetQuick PCR Purification Spin columns (Genomed, Lohne, Germany) as per manufacturer’s instructions. Sequencing was performed using Sequitherm Excel II sequencing kits (Epicenter, Madison, WI, USA) and IRD-labeled M13 primers (MWG Biotech, Ebersberg, Germany) on a LiCOR 4200 automated sequencer (Lincoln NE, USA), or by using Big-Dye Terminators (BDT) version 3.1 (Applied Biosystems, CA, USA) on an ABI PRISM 3100 Genetic Analyser (Applied Biosystems). Direct BAC sequencing was used to develop the sequence at the 5’ and 3’ ends of the gene and within intron 14 as previously described [34].

Polymorphism Identification and Genotyping

Polymorphisms were identified by comparing the chromatograms of pooled sequence with that of a single animal, and confirmed by genotyping of animals within the pool. Restriction fragment length polymorphism (RFLP) or sequencing (when no restriction enzymes were available) was used to genotype individuals (Table 1).

Bioinformatics and Statistical Analyses

Sequencer (Gene Codes, MI USA) was used to visualise chromatograms. We determined equine ACE haplotypes to allow for association testing in the future using Phase 2.0.2 [35] as we lacked parent-progeny combinations to derive haplotypes empirically. Cross species amino acid similarity and identity was scored using the program MatGAT [36]. Hydropathy analysis of predicted amino-acid sequence was carried out using the Kyte and Doolittle scale with a window size of 11 residues [37]. The transmembrane domain was identified using the Statistical Analysis of Protein Sequences (SAPS) package [38] and SignalP used to predict signal peptide cleavage sites [39]. Orthologs of the equine ACE gene were identified with Ensembl and used to create conservation plots with VISTA tools, using the LAGAN global alignment [40]. Multiple sequence alignments were performed using ClustalW2 with default settings. Phylogenetic analysis was performed with PhyML using the default settings on the horse, dog, cat, dolphin, cow, rabbit, elephant, human, chimpanzee, rat and mouse orthologous ACE genes [EnSembl: ENSE- CAG000000012990, ENSSCAF00000012998, EN- SFCAG00000002078, ENSTTRG0000001667, EN- SBTAG00000024950, ENSOUCUT0000001559,
Results and Discussion

identify transcription factor binding sites [40,42–47]. MatInspector, Alibaba2, MAPPER and rVISTA were all used to outgroup [41]. The internet based tools ConSite, TFSearch, SMUST00000001963 respectively, using the elephant as an outgroup [41]. The internet based tools ConSite, TFSearch, MatInspector, Alibaba2, MAPPER and rVISTA were all used to identify transcription factor binding sites [40,42–47].

Equine ACE Gene and Predicted Amino Acid Sequence

The genomic sequence of the equine ACE gene [GenBank:JX227848] was derived with the exception of the central regions of 3 large introns (18, 20 and 23), which comprise 3.54% of the total predicted gene sequence when compared to the horse genome reference sequence (Figure 1). In agreement with the structure of other mammalian ACE genes, the equine gene consisted of 26 exons. The cDNA showed a high level of conservation, both in sequence and exon size, between the horse and rabbit, human, rat and mouse genes, which were 87, 86, 84 and 84% homologous respectively. Sequencing the cDNA confirmed that exon 13 is not transcribed in the somatic form of equine ACE (sourced from leukocytes). The exons ranged from 75% (exon 26) to 95% (exon 9) similarity between horse and human (Table 2), with only exons 1, 13 and 26 differing in size from the human gene. The additional nucleotides in exons 1 and 15 in the horse compared to the human form part of the signal peptides and are thus cleaved from the mature enzymes. Similarly, the nucleotides increasing the size of exon 26 in the horse compared to the human are found in the 3' untranslated region (UTR). A high degree of homology was observed between exons 4 to 11 (region 1, Figure 1) and 17 to 24 (region 2, Figure 1), consistent with evidence for ancestral duplication of this gene [25]. In contrast to the exons, the introns showed little conservation across species in sequence, although the sizes were roughly similar, with the exceptions of intron 12 and 16. Intron 12 was 81% homologous to human intron 12, allowing identification of the putative testicular ACE promoter elements, while the first half of intron 16 was up to 77% similar to the homologous human intron. Intron 14 was found to contain an equine repetitive element-2 (ERE-2) [48].

The first 360 bp of the 5' region upstream of exon 1 was sequenced. This included some of the equine sACE promoter region, which was aligned with the human [GenBank:AF118569.1] and mouse [GenBank:M34433] promoter sequences (Figure S1, part a). Only the 130 bp directly upstream of the putative transcriptional start site (TSS) showed a significant homology of 75% and 88% to the human and mouse promoters respectively. This sequence corresponds to the 132 bp known to confer promoter activity to the human gene, and includes three potential SP1 binding sites homologous to the equivalent functional sites in the human [49]. No consensus CCAAT element was identified, which is in agreement with the comparison of human and mouse genes [25,50]. A similar comparison was performed for the tACE promoter (Figure S1, part b). The testicular promoter TTATT sequence was 15 bp upstream of the predicted TSS while the cAMP-responsive element binding site was conserved between the horse, human and mouse sequences [51].

The equine somatic ACE sequence including start and stop codons comprised 3942 nucleotides (1313 amino acids). The testicular enzyme comprises 737 amino acids, 72 of which are unique to this particular isoform. The aa sequence showed highest similarity to the cow and the least to the rat (Table 3). The signal peptide cleavage sites were predicted to be between residues Ala36 and Leu57 in sACE and Ser28 and Gln29 in tACE, making the mature enzymes 1277 and 709 aa long. Two metalloprotease active sites were predicted to be present in exons 8 and 21. The first showed the consensus sequence (H-E-M-G-H) present in other species. The second site differs by one residue, containing an isoleucine instead of a methionine in the central position, although the significance (if any) of this is unknown. Hydropathy analysis indicated that the C-terminal segment most probably anchors the enzyme to the cell membrane, similar to the pig [52], and amino acids 1268–1283 were identified as the most likely transmembrane segment.

Phylogenetic analysis of the cDNA was carried out on the horse, dog, cat, dolphin, cow, rabbit, elephant, human, chimpanzee, rat and mouse orthologous ACE genes. The resultant ML consensus tree (Figure 2) showed good agreement with the current accepted phylogenetic relationships of placental mammals, although bootstrapping indicated lower confidence in the position of the primates and outgroup Proboscidea. The Tasmanian devil sequence was removed from the original phylogenetic analysis, as it could not be placed within the tree with confidence. This was not unexpected as marsupials are estimated to have diverged from placental mammals between 185–225 million years ago [53].

Comparative Cross Species Analysis of Intron 16

Conservation across the orthologous horse, dog, rabbit, elephant, human, mouse and Tasmanian devil ACE genes is illustrated in Figures S2 and 3. Introns 12, 16 and 20 had conservation levels nearing that of the exons. Intron 12 is known to contain the testicular ACE promoter, and as such a high level of sequence similarity is expected in this area. The conservation
Comparative genomics has identified many non-protein-coding sites across intron 20 is not apparent in the rodent or marsupial, and no further analysis was carried out in this region. Further conservation analysis using rankVISTA across the entire gene confirmed that intron 12 and 16 were the only non-coding regions to show significant conservation across all species examined except the marsupial, so intron 16 became the focus of further analysis.

Figure 3a shows the conservation plot between the horse and the dog, human, elephant, rabbit and mouse intron 16 sequences. These species were included as representative species of the more diverse major clades of placental mammals (Perissodactyla, Carnivora, Primates, Proboscidea, Lagomorpha, and Rodentia; Figure 2). The marsupial was dropped from the analysis because it did not show conservation in this region. Within this intron lay a 380 bp region that was 77% identical between the horse and human (calculated in rankVISTA), while random genomic sequences are expected to be around 33% identical [42]. Within this conserved sequence, an 18 bp sequence was found to be identical across the six species investigated (Figure 3b).

Since this small sequence has been conserved over a range of placental mammals that are evolutionarily separated by around 103 million years [53], it is possibly a conserved nuclear element (CNE). Its absence from the marsupial places its origin between 103–185 million years ago. The conserved nuclear element could potentially be a cis-regulatory element. These elements are known to have similar or even lower conservation than promoter regions but still show a significant level of conservation due to selection pressure to maintain their activity, by elimination of mutations that disrupt function [54]. More divergence in these sequences (compared to protein coding and promoter sequences) is tolerated as there is some flexibility in the binding sites within cis-regulatory elements [54].

Comparative genomics has identified many non-protein-coding conserved sites in mammalian sequences by cross species comparison. Although few of these sites have defined function, it is thought they play important biological roles [55]. A study by Shen et al. found that most cis-regulatory elements identified in the mouse genome functioned by modifying transcription [55], and this effect is often tissue specific. Comparison of the homology of 320 kb genomic DNA surrounding the human, mouse and chicken stem cell leukaemia genes was successfully used to identify known and new enhancer elements [56]. Similarly, a novel sequence was found to regulate the interleukin-4, -5 and -13 genes across a number of mammalian species by comparing the sequence surrounding these genes [57]. Other examples are reviewed by Nobrega and Pennacchio [58], who also recognised that similar strategies will probably identify many more gene regulatory elements in the human genome.

A search for transcription factor (TF) binding sites across ACE intron 16 identified three TFs that were predicted to bind to the 18 bp CNE region by at least four of the five programs used (ConSite, TFSsearch, MAPPER, Alibaba2 and rVISTA). These were octamer-binding protein-1 (Oct-1), and hepatocyte nuclear factors 3-beta (HNF-3β, also known as FoxA2) and homologue-1 (HFH-1, also known as FoxQ1). The Fox family of factors are expressed in a range of tissues and have a wide ranging number of roles, including contributing to embryonic development, cell cycle regulation, cellular signalling and regulation of tissue specific gene expression [59,60]. In particular, HNF-3β is essential for notochord formation in embryonic development and regulates cell specific transcription in hepatocytes, intestinal, oesophageal, stomach and pancreatic epithelium [61–64]. Oct-1 is a ubiquitously expressed transcriptional regulator that is essential for embryonic survival and normal erythropoiesis [65,66]. Oct-1 is also thought to be a sensor for metabolic stress, recognising cellular stress and modulating gene expression in response [65,67]. Further functional studies are required to verify whether any of these transcription factors interact with the CNE region in intron 16 of the ACE gene.

We combined comparative sequence analysis with a search of known transcription factor binding sites to underpin the likelihood of identifying a functional non-coding regulatory region [68]. When tested for more than 100 wide ranging functional binding
Table 2. Comparison of the human and horse ACE exons and intron sequences.

| Number | Exon Size (bp) | Intron Size (bp) |
|--------|----------------|------------------|
|        | ECA            | HSA             | ECA            | HSA             |
| 1      | 292            | 271             | 81.5           | 616             | 587             |
| 2      | 168            | 168             | 85.1           | 612             | 908             |
| 3      | 94             | 94              | 93.6           | 668             | 668             |
| 4      | 144            | 144             | 87.5           | 441             | 441             |
| 5      | 192            | 192             | 89.1           | 577             | 562             |
| 6      | 98             | 98              | 89.8           | 399             | 377             |
| 7      | 173            | 173             | 86.1           | 331             | 727             |
| 8      | 224            | 224             | 85.3           | 345             | 339             |
| 9      | 145            | 145             | 95.2           | 302             | 286             |
| 10     | 99             | 99              | 92.9           | 290             | 290             |
| 11     | 123            | 123             | 93.5           | 364             | 358             |
| 12     | 212            | 212             | 85.8           | 291             | 299             |
| 13     | 242            | 228             | 77.7           | 126             | 169             |
| 14     | 137            | 137             | 92.7           | 1083            | 1184            |
| 15     | 159            | 159             | 84.3           | 278             | 270             |
| 16     | 88             | 88              | 83.0           | 2095            | 1574*            |
| 17     | 144            | 144             | 88.9           | 149             | 149             |
| 18     | 161            | 205             | >1584          | 1821            |
| 19     | 192            | 192             | 90.1           | >1513           | 2054            |
| 20     | 173            | 173             | 87.9           | >1513           | 2054            |
| 21     | 224            | 224             | 87.5           | 311             | 262             |
| 22     | 145            | 145             | 86.9           | 334             | 305             |
| 23     | 99             | 99              | 91.9           | >2105           | 1923            |
| 24     | 123            | 123             | 93.5           | 360             | 281             |
| 25     | 188            | 188             | 87.8           | 175             | 151             |
| 26     | 521            | 481             | 75.0           | -               | -               |

Exons and introns are compared by size, and exons by sequence identity. Cross species identity between the introns was generally too low to align and thus calculate. Introns 18, 20 and 23 were not fully sequenced. Exons shown in italics differ in size between the two species.

*Human intron 16 varies in length from 1574–1859 bp, depending on the presence or absence of the I/D polymorphism.

do:10.1371/journal.pone.0055434.t002

Table 3. Homology of equine ACE amino acid sequence to 5 different species.

| Species          | Genbank Accession No. | Identity (%) | Similarity (%) |
|------------------|-----------------------|--------------|----------------|
| Cow (Bos Taurus) | 1919242A              | 87.1         | 93.4           |
| Human (Homo sapiens) | NP_000780         | 86.7         | 93.5           |
| Chimpanzee (Pan troglodytes) | AAG31358        | 86.5         | 93.2           |
| Mouse (Mus musculus) | P09470            | 83.2         | 92.5           |
| Rat (Rattus norvegicus) | AAP80808         | 82.9         | 92.1           |

do:10.1371/journal.pone.0055434.t003

sites, the ConSite service retained around 70–80% of validated sites, whilst eliminating a number of false positives [69]. A similar process has also been used to identify regulatory modules across different species of *Drosophila* [70,71]. We consider it possible that the 18 bp CNE in intron 16 encodes a transcription factor binding site, although further functional analysis is needed to determine which of the factors might have an effect on ACE expression; in which tissues, and to what extent this modifies expression of the gene.

Polymorphism and Haplotype Analysis

Over 10 kb of sequence derived from 35 individuals and including 73% of the cDNA was screened for polymorphisms, resulting in the identification of 16 sequence changes (Figure 1, Tables 1 and S1). Eleven single nucleotide polymorphisms (SNPs) were identified in non-coding sequence and four were observed in coding sequence, including three that were silent, and one causing a conservative amino acid change (p.Arg1290His). This polymorphism is predicted to be within the intracellular region of the protein and as such is unlikely to play a role in circulating enzyme function. Additionally, the rat possesses a histidine in this position, so we expect that this exchange has no major effect on gene function. No I/D polymorphism was detected in intron 16, while a variable length poly-A stretch was identified in intron 14 associated with the equine repetitive element.

The use of pooled DNA and targeting the screen to coding regions decreased the number of polymorphisms detected in this study compared to those discovered in a similar study of the human ACE gene [72]. However, the proportion of common polymorphisms detected was comparable between the two species (63% of equine SNPs were found in more than one individual, compared to 67% in humans), confirming the utility of pooled DNA for detection of medium frequency common SNPs. An in-depth analysis of variation in the canine ACE gene in 100 individual dogs of different breeds identified 81 variants, including 4 in exons [73]. Although our study found fewer intronic variants, we only covered half the gene in our scan, in addition to using pooled DNA. Furthermore, we found 4 variants in coding sequence and another in the 3'UTR of exon 26, which was comparable to the number found in the higher coverage canine scan.

Nine SNPs were found in more than one animal (allocated SNPs 1–9, Table 1) and were genotyped across the panel of four different horse breeds (Standardbreds, SB; Arabians, AR; Thoroughbreds, TB; and heavy horses, HH). The nine equine SNPs were resolved into nine likely haplotypes (Figure 4a). From the 80 possible representations, one haplotype (H1) was represented 47 times, two (H6 and H7) 7 times, one (H2) 6 times, one (H9) 5 times, two (H5 and H8) 3 times and only two haplotypes (H3 and H4) were unique, in the HH and SB populations, respectively.

The haplotype containing the most common allele of each SNP occurred most frequently in our population. However, with the exception of haplotype 6 (which was seen in all breeds except the Thoroughbred), the other haplotypes appeared to occur either in the light horses (TB and AR, haplotypes 7, 8 and 9) or the heavy horses (SB and HH, haplotypes 2, 3 and 4, Figure 4b). Due to
the small numbers of horses screened it is not possible to identify whether the differences in haplotype distribution between breeds is significant and related to selective breeding for differential performance, or due to founder effects and closed studbooks. Haplotypes 4, 5, 8 and 9 contained one polymorphism within intron 16, while haplotype 6 contained two in complete linkage disequilibrium. Although none of these SNPs were within the 17 bp conserved sequence, SNPs 3 and 4 (in haplotypes 5, 6, 8 and 9) were within the first part of the intron that showed high homology to the corresponding human intron.

Although none of the equine SNPs identified coincide with the predicted TF binding site, or due to founder effects and closed studbooks. Haplotypes 4, 5, 8 and 9 contained one polymorphism within intron 16, while haplotype 6 contained two in complete linkage disequilibrium. Although none of these SNPs were within the 17 bp conserved sequence, SNPs 3 and 4 (in haplotypes 5, 6, 8 and 9) were within the first part of the intron that showed high homology to the corresponding human intron.

Although none of the equine SNPs identified coincide with the predicted TF binding site, at least one human SNP is known to occur at base 10 of the 18 bp CNE (Figure 3b). This A to G transition (rs4338) is predicted to reduce affinity for this site and thus prevent binding of all three of the identified transcription factors, potentially altering gene transcription. This SNP is not in complete linkage disequilibrium (LD) with the I/D polymorphism, and is rare in the human population, with a minor allele frequency of approximately 0.017% [74]. The G allele is associated with the deletion Alu allele and higher circulating enzyme levels [72].

This SNP in particular warrants specific testing in association studies in humans. Other SNPs in close proximity, rs4334, rs4336 and rs4337 (287, 96 and 94 bp upstream of rs4338 respectively) have been observed to be in complete LD with the I/D polymorphism [72] and occur in the highly conserved region just upstream of the putative binding site. The observed variation in human circulating ACE levels attributed to the I/D polymorphism may actually be due to the action of one of these three SNPs. SNPs
Multi-Species Comparison of the Equine ACE Gene

Base genome: Horse  Chromosome: HORSE_ACE  8,940-12,400

3a

Dog

Human

Elephant

Rabbit

Mouse

3b

Horse bp: 10444

Horse  CCAACTCCACATAGGAAAAATGAAAATAAACAGGAGGAGAA
Dog  CCAACTCCACATAGGAAAAATGAAAATAAACAGGAGGAGAG
Human  CCAACTCCACATAGGAAAAATGAAAATAAACAGGAGGAGAG
Elephant  CCAACTCCACATAGGAAAAATGAAAATAAACAGGAGGAGAG
Rabbit  CCAACTCCACATAGGAAAAATGAAAATAAACAGGAGGAGAG
Mouse  CCAACTCCACATAGGAAAAATGAAAATAAACAGGAGGAGAG

3c

rs4338

PLOS ONE | www.plosone.org 7 February 2013 | Volume 8 | Issue 2 | e55434
Figure 3. Sequence conservation across intron 16 of multi species ACE genes. (a) Plot of conservation spanning exons 15 to 18 between the developed equine ACE sequence and the reference sequences from the dog, human, elephant, rabbit and mouse. Pink coloured regions are ≥70% conserved between the horse and query sequence; while the dark blue regions are annotated exons. Pink conserved peaks are clearly visible in intron 16. The positions of common equine SNPs are marked with black arrows, and the human I/D variant is noted in blue. The human intron sequence analysed here was of the deletion allele. (b) Aligned intron sequence from six species showing the conserved nucleotides highlighted in yellow. The position of SNP rs4338 that occurs in the middle of the human CNE is highlighted in green. (c) Results of the rVISTA scan showing likely motif binding sites (green bars) for the three identified transcription factors.

Figure 4. Haplotypes identified in the equine ACE gene. (a) Haplotype representation, with blue boxes showing major alleles, red boxes minor alleles and the yellow box shows the third allele of SNP 6. (b) Haplotype distributions across the whole population and divided into the different breeds examined: Thoroughbred (TB), Arabian (AR), Standardbred (SB) and heavy horses (HH, including Clydesdales and Shires).

Conclusions
We have performed an extensive study of the sequence and structure of the equine ACE gene, and identified common haplotypes of the gene across a diverse cohort of breeds. We identified a conserved non coding element within intron 16 that is...
shared across representatives of the major placental mammalian lineages. It provides a new focus for the identification of functional variants within the ACE gene that affect enzyme levels and biological performance. Soubrier and colleagues [84] noted that since ACE has been extensively and systemically sequenced it is likely that all the functional variants have been detected, but the identification is impeded by their almost complete LD with the I/D polymorphism. Further study of the SNPs recognised in this study (both within the horse and human) may uncover the functional variant that has previously eluded researchers.

Supporting Information

Figure S1  Alignment of the horse, human and mouse ACE gene promoter sequences. Figure S1 part a shows the alignment of the somatic ACE promoters with the TATATA box highlighted in yellow. Putative SP1 binding sites known to be functional in the human are indicated in green. Part b shows the alignment of the intronic testicular ACE promoters. The TTATTA sequence is highlighted in yellow and the predicted CAMP-responsive element binding site in green. (DOCX)

Figure S2  Full multi-species alignment of the ACE gene. The alignment shows the conservation between the developed equine ACE sequence with the dog, human, elephant, rabbit, mouse and Tasmanian devil orthologous ACE gene sequences [EnSembl: ENSCAFG000000012998, ENST000000298066, ENSLAFG00000006295, ENSOCUT0000001559, ENSMUST00000019163 and ENSSHAT0000012503 respectively].

Regions that are coloured pink are $\geq 70\%$ conserved between the reference and query sequences, and the dark blue regions are annotated exons. Exon 13, which is not transcribed into the sACE protein, is not annotated. Pink conserved peaks are clearly visible in introns 12 and 16 (which are labelled) across most species, but not in other introns. (DOCX)

Table S1  Primers for characterization of the equine ACE gene. All primers were used for sequencing of PCR product with the exceptions of Aceex1rev, AceI14rev Aceex26for, which were used for direct BAC sequencing. Primer pairs also used for screening the DNA pools are marked (*). (DOC)

Acknowledgments

Francois Pumi, INRA, Centre de Recherche de Jouy, Laboratoire de Génétique biochimique et de Cytogénétique, Jouy-en-Josas, France, provided the BAC clone containing the gene. Thanks for technical support are due to Marilyn Jones, Gina Attard and Johanna Lang-Davis. We are very grateful to Trent Haymen, Greg Hogan, Byron Biffin and a racing stable veterinarian and owners for supplying equine blood samples.

Author Contributions

Collected samples: NAH. Conceived and designed the experiments: NAH IT HR. Performed the experiments: NAH. Analyzed the data: NAH. Contributed reagents/materials/analysis tools: NAH IT HR. Wrote the paper: NAH IT HR.

References

1. Cole J, Ettery D, Bernstein KE (2000) Insights derived from ACE knockout mice. J Renin Angiotensin Aldosterone Syst 1: 157–141.
2. Cambien F, Costerousse O, Tiet C, Poirier O, Lecerf L, et al. (1994) Plasma gene and liver polymorphism of angiotensin-converting enzyme in relation to myocardial infarction. Circulation 90: 669–676.
3. Rigat B, Hubert C, Alencar-Gelas F, Cambien F, Corvol P, et al. (1990) An Insertion/Deletion polymorphism in the angiotensin I-converting enzyme gene accounting for half the variance of serum enzyme levels. J Clin Invest 86: 1343–1346.
4. Cambien F, Poirier O, Lecerf L, Evans A, Cambou J, et al. (1992) Deletion polymorphism in the gene for angiotensin-converting enzyme is a potent risk factor for myocardial infarction. Nature 359: 641–644.
5. Evans A, Poirier O, Kee F, Lecerf L, McCrum E, et al. (1994) Polymorphisms of the angiotensin-converting enzyme gene in subjects who die from coronary heart disease. Q J Med 87: 211–214.
6. Fujisawa T, Ikegami H, Kawaguchi Y, Hamada Y, Ueda H, et al. (1998) Meta-analysis of association of insertion/deletion polymorphism of angiotensin I-converting enzyme gene with diabetic nephropathy and retinopathy. Diabetologia 41: 47–53.
7. Castellon R, Handl HK (2007) Denaturing the ACE polymorphism: from genetics to biology. Curr Pharm Des 13: 1191–1198.
8. Montgomery HE, Marshall R, Hemingway H, Myers S, Clarkson P, et al. (1998) Human gene for physical performance. Nature 393: 221–222.
9. Gavagay G, Yu B, Hambly B, Boston T, Halan A, et al. (1998) Elite endurance athletes and the ACE I allele—the role of genes in athletic performance. Hum Genet 103: 48–50.
10. Alcaraz K, Terrados N, Ortolano R, Iglesias-Cubero G, Reguero JR, et al. (2000) Genetic variation in the renin-angiotensin system and athletic performance. Eur J Appl Physiol 82: 117–120.
11. Myerson S, Hemingway H, Martin J, Humphries S, et al. (1999) Human angiotensin I-converting enzyme gene and endurance performance. J Appl Physiol 87: 1313–1316.
12. Nazarov I, Woods D, Montgomery H, Shnider C, Kazakov V, et al. (2003) The angiotensin converting enzyme I/D polymorphism in Russian athletes. Eur J Hum Genet 9: 797–801.
13. Folland J, Leach B, Little T, Howells K, Myerson S, et al. (2000) Angiotensin-converting enzyme genotype affects the response of human skeletal muscle to functional overload. Exp Physiol 85: 573–579.
14. Williams A, Rayson M, Jubb M, World M, Woods D, et al. (2000) The ACE gene and muscle performance. Nature 403: 614.
15. Sonna L, Sharp M, Knapik J, Sullivan M, Angel K, et al. (2001) Angiotensin-converting enzyme genotype and physical performance during US army basic training. J Appl Physiol 91: 1355–1363.
16. Karjalainen J, Kujsa M, Stolt A, Mantysaari M, Vitasalo A, et al. (1999) Angiotensinogen gene M235T polymorphism predicts left ventricular hypertrophy in endurance athletes. J Am Coll Cardiol 34: 494–499.
17. Taylor R, Marnote C, Fallon K, van Bockxmeer F (1999) Elite athletes and the gene for angiotensin-converting enzyme. J Appl Physiol 87: 1033–1037.
18. Rankinen T, Wollarth B, Simmeone J, Maer-Lenz D, Kraaumma R, et al. (2000) No association between the angiotensin-converting enzyme ID polymorphism and elite endurance athlete status. J Appl Physiol 88: 1571–1573.
19. Tiet C, Rigat B, Viviikis S, Breda G, Corvel P, et al. (1992) Evidence, from combined segregation and linkage analysis that a variant of the angiotensin I-converting enzyme (ACE) gene controls plasma ACE levels. Am J Hum Genet 51: 197–205.
20. Suetoro T, Mori T, Inoue M, Kunom Y, Breda Y, et al. (2004) Sequeletion and linkage analysis of a variant of the angiotensin I-converting enzyme (ACE) gene controls plasma ACE levels. Am J Hum Genet 51: 197–205.
21. Kaelin L, Gupta A, Breda G, Corvel P, et al. (1992) Evidence, from combined segregation and linkage analysis that a variant of the angiotensin I-converting enzyme (ACE) gene controls plasma ACE levels. Am J Hum Genet 51: 197–205.
22. Suetoro T, Mori T, Inoue M, Kunom Y, Breda Y, et al. (2004) Sequeletion and linkage analysis of a variant of the angiotensin I-converting enzyme (ACE) gene controls plasma ACE levels. Am J Hum Genet 51: 197–205.
23. Resotto N, Pontremoli R, de Ferrari G, Razvadzal R (1999) Intronic 16 insertion of the angiotensin converting enzyme gene and transcriptional regulation. Nephrol Dial Transplant 14: 868–871.
24. McKenzie CA, Abecasis GR, Keavney B, Forrester T, Ratcliffe PJ, et al. (2001) Trans-ethnic fine mapping of a quantitative trait locus for circulating angiotensin I-converting enzyme (ACE) Hum Mol Genet 10: 1077–1084.
25. Villard E, Tiet C, Viviikis S, Rakotowsa R, Cambien F, et al. (1996) Identification of new polymorphisms of the angiotensin I-converting enzyme (ACE) gene, and study of their relationship to plasma ACE levels by two-QTL segregation-linkage analysis. Am J Hum Genet 58: 1250–1278.
26. Zha Y, Bouzekri N, Southam L, Cooper RS, Adeyemo A, et al. (2001) Linkage and association analysis of angiotensin I-converting enzyme (ACE) gene polymorphisms with ACE concentration and blood pressure. Am J Hum Genet 68: 1139–1148.
27. Hubert C, Houst A, Corvel P, Soubrier F (1991) Structure of the angiotensin I-converting enzyme gene. Two alternative promoters correspond to evolutionary steps of a duplicated gene. J Biol Chem 266: 15377–15383.
28. Bandin B (2002) New aspects on angiotensin-converting enzyme: from gene to disease. Clin Chim Lab Med 40: 236–263.
29. Bellent V, Michael A, Bonnefoy C, Chauvet M-T, Corvel P (1995) Cell surface localization of proteolysis of human endothelial angiotensin I-converting enzyme. Effect of the amino-terminal domain in the solubilization process. J Biol Chem 270: 28062–28069.
25. Woodman Z, Oppong S, Cook S, Hooper N, Schwager S, et al. (2008) Shredding of somatic angiotensin-converting enzyme (ACE) is inefficient compared with tests ACE despite cleavage at identical stalk sites. Biochem J 347: 711–718.

26. Souther F, Alhene-Gelas F, Hubert C, Allegri J, John M, et al. (1988) Two putative active sites in human angiotensin I-converting enzyme revealed by molecular cloning. Proc Natl Acad Sci U S A 85: 9396–9390.

27. Coomer RPC, Forhead AJ, Bathe AP, Head MJ (2003) Plasma angiotensin-converting enzyme (ACE) concentration in Thoroughbred racehorses. Equine Vet J 35: 96–98.

28. Godard S, Schibler I, Oustry A, Criqui EP, Guerin G (1998) Construction of a horse BAC library and cytogénnetical assignment of 20 type I and type II markers. Mammalian Genome 9: 633–637.

29. Montgomery G, Sue J (1990) Expression from DNA from sheep white blood cells. NZ J Agric Res 33: 437–441.

30. Oetting W, Lee H, Flanders G, Sellers T, et al. (1995) Linkage analysis with multiplexed short tandem repeat polymorphisms using infrared fluorescence and M13 tagged primers. Genomics 30: 450–458.

31. Cavanagh JAL, Brinton S, Lalmay J, Semeniuk D, Marino EM, Zhou Y, Overbeek P, et al. (1997) Characterization of the bovine age gene: genomic structure and physical linkage mapping. Ani Genet 36: 435–462.

32. Stephens M, Smith N, Doucet P (2001) A new statistical model for haplotype reconstruction from population data. Am J Hum Genet 68: 978–989.

33. Campanella JJ, Mittlecke S, Smallley J (2003) MapAT: An application that generates similarity/identity matrices using protein or DNA sequences. BMC Bioinformatics 4: 1–6.

34. Kyte J, Doolittle R (1982) A simple method for displaying the hydropathic character of a protein. J Mol Biol 157: 105–132.

35. Rausa FM, Galarneau L, Belanger L, Costa RH (1999) The nuclear receptor foprotein transcription factor expression is coexpressed with its target gene ACE in the developing murine liver, intestine and pancreas. Mech Dev 89: 185–188.

36. Kramers C, Danilov SM, Deinum J, Balyasnikova IV, Scharenborg N, et al. (2010) Comparative genomic analysis as a tool for identifying regulatory elements. Nat Rev Genet 5: 276–287.

37. Korsgren O, Hindley J, Kuhlen M, Petersen M (1998) Analyzing global DNA sequence Alignments of Arbitrary Length. Bioinformatics 14: 1046–1047.

38. Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, et al. (2010) New algorithms for statistical analysis of protein sequences. Proc Natl Acad Sci USA 89: 2002–2006.

39. Nielsen H, Engelbrecht J, Brunak S, von Heijne G (1997) Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. Protein Eng 10: 365–369.

40. Matsubara A, Aiba H, Mitani T, Morita K, Maruyama S, et al. (2002) Identification of conserved regulatory elements by comparative genome analysis. J Biol 2: 12–17.

41. Marinasuc VD, Kohnen IS, Riva A (2005) MAPPER: a search engine for the comparative sequence-based discovery of functional transcription factor binding sites in multiple genomes. BMC Bioinformatics 6: 79.

42. Luhr B, Bollard A, Mendoza L, Engstrom P, Jarzebski N, et al. (2003) Comparative genomic analysis as a tool for identifying regulatory elements. Nat Rev Genet 5: 276–287.

43. Oetting W, Lee H, Flanders G, Sellers T, et al. (1995) Linkage with equine microsatellite loci. Mamm Genome 10: 140–144.

44. Grabe N (2002) AliBaba2: context specific identification of transcription factor binding sites in multiple genomes. BMC Bioinformatics 4: 1–6.

45. Heinemeyer T, Wingender E, Reuter I, Hermjakob H, Kel AE, et al. (1998) MatInspector: new fast and versatile tools for detection of consensus matches in multiple genomes. BMC Bioinformatics 5: 129.

46. Thousand genomes consortium (2010) A map of human genome variation from population-scale sequencing. Nature 467: 1061–1073.

47. Rieder MJ, Taylor SL, Clark AG, Nickerson DA (1999) Sequence variation in the mouse HNF-3/hoX gene: genomic structure and physical characterization of the mouse age gene: genomic structure and functional transcription factor binding-site clusters in Drosophila. BMC Bioinformatics 5: 129.

48. Korsgren O, Hindley J, Kuhlen M, Petersen M (1998) Analyzing global DNA sequence Alignments of Arbitrary Length. Bioinformatics 14: 1046–1047.

49. Thousand genomes consortium (2010) A map of human genome variation from population scale sequencing. Nature 467: 1061–1073.

50. Korsgren O, Hindley J, Kuhlen M, Petersen M (1998) Analyzing global DNA sequence Alignments of Arbitrary Length. Bioinformatics 14: 1046–1047.

51. Korsgren O, Hindley J, Kuhlen M, Petersen M (1998) Analyzing global DNA sequence Alignments of Arbitrary Length. Bioinformatics 14: 1046–1047.

52. Korsgren O, Hindley J, Kuhlen M, Petersen M (1998) Analyzing global DNA sequence Alignments of Arbitrary Length. Bioinformatics 14: 1046–1047.

53. Korsgren O, Hindley J, Kuhlen M, Petersen M (1998) Analyzing global DNA sequence Alignments of Arbitrary Length. Bioinformatics 14: 1046–1047.