Analysis of Prostate-Specific Antigen Transcripts in Chimpanzees, Cynomolgus Monkeys, Baboons, and African Green Monkeys

James N. Mubiru1*, Alice S. Yang2, Christian Olsen3, Sudhir Nayak4, Carolina B. Livi5, Edward J. Dick Jr.1, Michael Owston1, Magdalena Garcia-Forey1, Robert E. Shade1, Jeffrey Rogers6

1 Southwest National Primate Research Center, Texas Biomedical Research Institute, San Antonio, Texas, United States of America, 2 Department of Biology, St. Mary’s University, San Antonio, Texas, United States of America, 3 Biomatters Inc., Newark, New Jersey, United States of America, 4 Department of Biology, The College of New Jersey, Ewing, New Jersey, United States of America, 5 Department of Molecular Medicine and Institute of Biotechnology, University of Texas Health Science Center, San Antonio, Texas, United States of America, 6 Human Genome Sequencing Center, Baylor College of Medicine, Houston, Texas, United States of America

Abstract

The function of prostate-specific antigen (PSA) is to liquefy the semen coagulum so that the released sperm can fuse with the ovum. Fifteen spliced variants of the PSA gene have been reported in humans, but little is known about alternative splicing in nonhuman primates. Positive selection has been reported in sex- and reproductive-related genes from sea urchins to Drosophila to humans; however, there are few studies of adaptive evolution of the PSA gene. Here, using polymerase chain reaction (PCR) product cloning and sequencing, we study PSA transcript variant heterogeneity in the prostates of chimpanzees (Pan troglodytes), cynomolgus monkeys (Macaca fascicularis), baboons (Papio hamadryas anubis), and African green monkeys (Chlorocebus aethiops). Six PSA variants were identified in the chimpanzee prostate, but only two variants were found in cynomolgus monkeys, baboons, and African green monkeys. In the chimpanzee the full-length transcript is expressed at the same magnitude as the transcripts that retain intron 3. We have found previously unidentified splice variants of the PSA gene, some of which might be linked to disease conditions. Selection on the PSA gene was studied in 11 primate species by computational methods using the sequences reported here for African green monkey, cynomolgus monkey, baboon, and chimpanzee and other sequences available in public databases. A codon-based analysis (dN/dS) of the PSA gene identified potential adaptive evolution at five residue sites (Arg45, Lys70, Gln144, Pro189, and Thr203).

Citation: Mubiru JN, Yang AS, Olsen C, Nayak S, Livi CB, et al. (2014) Analysis of Prostate-Specific Antigen Transcripts in Chimpanzees, Cynomolgus Monkeys, Baboons, and African Green Monkeys. PLoS ONE 9(4): e94522. doi:10.1371/journal.pone.0094522

Editor: A. Aziz Abouleban, University of Oxford, United Kingdom

Received May 17, 2013; Accepted March 18, 2014; Published April 14, 2014

Copyright: © 2014 Mubiru et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: Dr. Mubiru was supported by the Office of the Director, National Institutes of Health under Award # 8K01OD010973-04. This investigation used resources that were supported by the Southwest National Primate Research Center grant P51 RR013986 from the National Center for Research Resources, National Institutes of Health and that are currently supported by the Office of Research Infrastructure Programs through P51 OD013986. Nonhuman primates were housed in facilities constructed with support from Research Facilities Improvement Programs Grants C06 RR015456 and C06 RR014578 from the National Center for Research Resources, National Institutes of Health. This work received computational support from Computational Systems Biology Core, funded by the National Institute on Minority Health and Health Disparities (G12MD007591) from the National Institutes of Health. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: Christian Olsen, one of the authors is employed by a commercial company Biomatters, Inc. as a Field Application Scientist. This does not alter the authors’ adherence to all the PLOS ONE policies on sharing data and materials.

* E-mail: jmubiru@txbiomed.org

Introduction

Prostate-specific antigen (PSA) is encoded by the kallikrein-3 (KLK3) gene, which belongs to the tissue kallikrein (KLK) gene family; the official name of the gene is KLK3, although it is commonly referred to as PSA [1,2]. The PSA gene is composed of five exons and four introns. Part of the first exon codes for a signal peptide that targets the protein for secretion [3,4]. In the PSA protein, residues His57, Asp102, and Ser195 have been reported to be important for proteolytic activity of PSA and other kallikreins, and they are commonly referred to as the catalytic triad [5,6]. PSA has an activation peptide of seven amino acids that is cleaved by KLK2 protein to generate enzymatically active PSA [7,8]. PSA is highly expressed by prostatic epithelial cells and is abundant in seminal plasma [9]. After ejaculation, semen forms a coagulum through the linkage of semenogelin proteins [10,11]. Then, the coagulum has to be liquefied so that the sperm can be released and fuse with the ovum. In humans, great apes, and Old World monkeys this is done by PSA [3,11–14].

Alternative splicing is a common mechanism in nature to enhance protein diversity, and it occurs in more than 90% of human genes [15,16]. Alternative splicing of the PSA gene produces at least 15 transcripts of 0.7–6.1 kb [17].

Positive selection has been reported in sex- and reproductive-related genes from sea urchins to Drosophila to humans [12,18,19]. However, there is scant data regarding the PSA gene despite its importance in reproductive success. Marques et al. (2012) recently provided evidence of adaptive evolution of the KLK3 gene toward an expanded enzyme spectrum that has an effect on the hydrolysis of semen coagulum [20]. In another study that analyzed the PSA gene, Clark and Swanson (2005) reported an overall dN/dS value of 0.38 for the PSA gene in human-chimpanzee pairwise comparisons, suggesting purifying selection [21].
The aims of this study are to investigate PSA transcript heterogeneity in four nonhuman primates commonly used in biomedical research and to study the mode of selection exerted on this gene in primates. Our previous data showed that cynomolgus monkeys have high serum PSA concentrations compared to baboons, which have low concentrations [22]. The chimpanzee was examined because of its close evolutionary relationship to humans.

Materials and Methods

Ethics Statement

We opportunistically sampled animals that presented for necropsy. Six cynomolgus monkeys (Macaca fascicularis) and baboons (Papio hamadryas anubis) and four chimpanzees (Pan troglodytes) were from the Southwest National Primate Research Center (SNPRC), located at the Texas Biomedical Research Institute in San Antonio, Texas. Three African green monkey (Chlorocebus aethiops) samples were from the Department of Pathology/Comparative Medicine, Wake Forest University of Health Sciences, Winston-Salem, North Carolina. The chimpanzees died of heart failure as a result of chronic cardiomyopathy. Three cynomolgus monkeys were removed from the colony and euthanized because of positive titers to simian retrovirus. One baboon with a sperm granuloma that tested positive for herpes was euthanized because of positive titers to simian retrovirus. One chimpanzee died of heart failure as a result of chronic cardiomyopathy. The Institutional Animal Care and Use Committee approved all procedures. Human prostate cDNAs were from cryopreserved tissues that were collected following radical prostatectomies under an Institutional Review Board approved protocol for the Department of Pathology at the University of Texas Health Science Center, San Antonio, and have been described previously [23].

Transcript Cloning

We extracted total RNA from prostate samples using Trizol reagent (Life Technologies, Carlsbad, CA), reverse-transcribed them, and performed polymerase chain reaction (PCR) analysis, as described previously [22]. Primer sequences were based on the cynomolgous monkey and chimpanzee sequences (GenBank accession numbers AY647976, NM_001009136) [12,24] and are shown in Tables 1 and 2. The forward primers were located in exon 1, and the reverse primers were located in exon 5 or in intron 3. We used primer sequences located in intron 3 because of reports indicating that some PSA variants contain sequences in the third intron [25]. To obtain the complete coding sequences, we cloned the PCR products into the pCR2.1-TOPO vector (Life Technologies, Carlsbad, CA) and sequenced them. Cloning using the two primer sets was done at least four times for each animal species. Each time 22 colonies were picked and cultured. Positive clones were characterized by restriction enzyme analysis according to their size on an agarose gel. One clone of each size was sequenced. We predicted the presence of a signal peptide to determine whether the transcripts could be secreted outside the cell using the Signal Peptide Prediction website (http://www.cbs.dtu.dk/services/SignalP/).

Real-Time PCR

A closer look at the transcripts indicated that they could be broadly grouped into those that used the sequences in intron 3 and those that did not. We performed real-time PCR to explore whether expression of transcripts that do not use intronic sequences takes precedence over those that do. We conducted quantitative reverse transcription (RT) PCR using TaqMan technology (Life Technologies, Carlsbad, CA). We designed primers and probes for the quantitation of the chimpanzee full-length and short-version transcripts using the IDTDNA SciTools software (Integrated DNA Technologies Inc., Coralville, IA); the sequences are shown in Table 2. TaqMan 18s ribosomal RNA probe and primers served as the internal control. All samples and standards were analyzed in triplicate. We conducted real-time fluorescence-based PCR using the ABI Prism 7900 real-time PCR thermal cycler (Applied Biosystems, Foster City, CA) under vendor-specified conditions. We determined the values of the unknown samples from the standard curves prepared in each assay and normalized the concentration of the PSA transcripts to 18s ribosomal RNA.

Analysis of Selection

We assessed the evolutionary selection acting on the PSA gene using 11 PSA gene sequences. Our group produced four sequences, one each for Pan troglodytes, Macaca fascicularis, Papio hamadryas anubis, and Chlorocebus aethiops; these sequences are reported in this paper. Eight other sequences—Macaca fascicularis (AY647976), Macaca mulatta (NM_001042776.1), Macaca fuscata (KC155626.1), Homo sapiens (NM_001648.2), Pan paniscus (DQ150478.1), Gorilla gorilla (KC155627.1), Pongo pygmaeus (DQ150480.1), and Nomascus leucogenys (UCSC genome Browser, Gibbon chr10)—were retrieved from public databases.

Protein sequences were aligned using the Multiple Sequence Comparison by Log-Expectation (MUSCLE) program, and codon-guided alignments were generated using Java Codon Delimited Alignment [JCoDA] [26,27]. Preliminary analysis of the number of synonymous substitutions per synonymous site (Ks) and the number of nonsynonymous substitutions per nonsynonymous site (Ka) was performed on all pairwise comparisons using the coding region and sliding window (window 20, jump 5) analysis in JCoDA. Although purifying selection dominated across all comparisons, multiple regions in the C-terminal region appeared to be under relaxed or positive selection or with dN/dS (ω) values greater than 2. To identify individual residues under positive selection, we used the Phylogenetic Analysis by Maximum Likelihood (PAML) software package, version 4.7, with the following comparisons: between null hypothesis M0 (one ratio for all branches) and hypothesis M3 (discrete), between hypotheses M1 (neutral) and M2 (selection), and between hypotheses M7 (B-distributed 0 to 1) and M8 (B-distributed and estimated ω). Sites under selection were identified using the Bayes empirical Bayes (BEV) approach, the naive empirical Bayes (NEB) approach, and the likelihood ratio test (LRT) in CODEML (part of the PAML package) [28–31]. Phylogenetic trees were constructed using neighbor-joining, maximum parsimony, maximum-likelihood, and unweighted pair group method with arithmetic mean (UPGMA) methods provided with the Phylib package, version 3.695 [32]. In each case, tree topology was tested using bootstrapping with 100 replicates.

We were cognizant of the reported chimeric nature of the KLK2/KLK3 genes in both Gorilla gorilla and Nomascus leucogenys. Marques et al. (2012) indicated that the gorilla and gibbon KLK3 gene is in fact a fusion of KLK3 and KLK2 (eKLK), in which the first four exons of the fused gene are orthologous to KLK3 and the last exon is more similar to KLK2 [20]; therefore in our analysis we excluded exon 5. However, it has also been reported that at the
protein level these genomic rearrangements in exon 5 account for only minor amino acid replacements and are not predicted to alter protein structure and function and that the "Kallikrein loop" is composed of an 11 amino acid insertion (95A–95K) that is specific for PSA [33].

When numbering the amino acid residues, we followed both the PSA protein numbering and also the convention used for serine proteases. The convention numbering system is based on chymotrypsin and denotes that the catalytic triad is composed of His57, Asp102, and Ser195 and that the "Kallikrein loop" is characteristic of the kallikrein (KLK) gene family, and PSA-2 is identical to a chimpanzee sequence available in the GenBank because no similar transcript was found in the NCBI GenBank.

The PSA-1 transcript has an open reading frame of 543 bases, which encodes a 180 amino acid protein, and has a predicted molecular weight of 28,839 Da (Figure 2). This sequence is orthologous to the human KLK3 transcript variant 1 (NM_001648.2). The cynomolgus monkey PSA-1 is identical to the one reported by Marshal et al. (2006) under GenBank accession number DQ150477 [21]. The African green monkey sequence is identical to sequences under GenBank accession numbers NM_001112745, DQ150485 [21], and EF676031 [22]. The chimpanzee sequence is identical to a sequence under GenBank accession number DQ150477 [21]. The African green monkey PSA-1 is novel, because no similar transcript was found in the National Center for Biotechnology Information (NCBI) GenBank.

The PSA-2 transcript has an open reading frame of 543 bases, which encodes a 180 amino acid protein, and has a molecular weight of 19,820 Da (Figure 2). The chimpanzee transcript is identical to a chimpanzee sequence available in the GenBank under accession number NM_001009136.1 [12] and is orthologous to gorilla and orangutan sequences numbers AY781395.1 and AY781391.1, respectively. It is 98% similar to a human sequence identified as PSA-RP2 reported by Heuze´-Vourc’h et al. (2001) (GenBank AJ310937, number CAC41631) [25]. The cynomolgus monkey, baboon, and African green monkey PSA-2 sequences are novel; no similar transcripts were found in the NCBI GenBank.

### Results

#### Transcript Cloning

We identified two PSA transcripts in the cynomolgus monkey, baboon, and African green monkey and six PSA transcripts in the chimpanzee prostate (Table 3; Figure 1). From all four species the same two main transcripts were identified: PSA-1 has the five exons characteristic of the kallikrein (KLK) gene family, and PSA-2 is missing exons 4 and 5 and uses sequences in intron 3. Because of intron retention, a unique 50-base sequence is added to exon 3. The PSA-1 transcript is derived from five exons with an open reading frame of 786 bases encoding a 261 amino acid protein and has a predicted molecular weight of 28,839 Da (Figure 2). This sequence is orthologous to the human KLK3 transcript variant 1 (NM_001648.2). The cynomolgus monkey PSA-1 is identical to the one reported by Marshal et al. (2006) under GenBank accession number DQ150477 [21]. The African green monkey PSA-1 is identical to the sequence under GenBank accession numbers NM_001112745, DQ150485 [21], and EF676031 [22]. The chimpanzee sequence is identical to a sequence under GenBank accession number DQ150477 [21]. The African green monkey PSA-1 is novel, because no similar transcript was found in the National Center for Biotechnology Information (NCBI) GenBank.

The PSA-2 transcript has an open reading frame of 543 bases, which encodes a 180 amino acid protein, and has a molecular weight of 19,820 Da (Figure 2). The chimpanzee transcript is identical to a chimpanzee sequence available in the GenBank under accession number NM_001009136.1 [12] and is orthologous to gorilla and orangutan sequences numbers AY781395.1 and AY781391.1, respectively. It is 98% similar to a human sequence identified as PSA-RP2 reported by Heuze´-Vourc’h et al. (2001) (GenBank AJ310937, number CAC41631) [25]. The cynomolgus monkey, baboon, and African green monkey PSA-2 sequences are novel; no similar transcripts were found in the NCBI GenBank.

### Table 1. Primer sequences used for cloning cynomolgus monkey, baboon, African green monkey, and chimpanzee PSA.

| Primer | Sequence 5′–3′ | Location of Primer | Product Size |
|--------|----------------|-------------------|--------------|
| PSA outer forward (A-07-2911) | 5′-CTGGACACGCTGTCACCACAT-3′ | Exon 1 | 1081 bases |
| PSA outer reverse (A-07-3585) | 5′-TGAGATGTCTCCAGGCACTGA-3′ | Exon 5 | |
| PSA inner forward (A-07-2910) | 5′-CTGTGTACCACATGTTGTC-3′ | Exon 1 | 811 bases |
| PSA inner reverse (A-07-2908) | 5′-GAAGTGTAGGCTGCTCAG-3′ | Exon 5 | |
| Chimpanzee full-length PSA cDNA cloning | | |
| Chimpanzee full-length reverse (12-3747) | 5′-CGGTGTTCCACTCTCGGGTCG-3′ | Intron 3 | 568 bases |
| Chimpanzee full-length forward (12-3746) | 5′-CTGTGTACCACATGTTGTC-3′ | Exon 1 | 808 bases |
| Chimpanzee short-length probe | 5′-GAGTTGATGGGGTGCTCAG-3′ | Exon 5 | |
| Chimpanzee short-length reverse | 5′-AGGGATGTTAAGGGGCTGCT-3′ | |
| Chimpanzee short-length forward | 5′-GAGTTGATGGGGTGCTCAG-3′ | |
| Chimpanzee short-length probe | 5′-AGGGATGTTAAGGGGCTGCT-3′ | |

### Table 2. Primer sequences used for real-time quantification of chimpanzee full-length and short-version PSA transcripts.

| Primer | Sequence 5′–3′ | Location of Primer | Product Size |
|--------|----------------|-------------------|--------------|
| Chimpanzee full-length forward | 5′-GCAGTTCCACCCTCAGAAG-3′ | Exon 4 | |
| Chimpanzee full-length reverse | 5′-GAAGCAACACATTACAGCAAG-3′ | Exon 4–exon 5 junction | |
| Chimpanzee short-length forward | 5′-CCCCAGATTACACCAGGCAAG-3′ | Exon 3 | |
| Chimpanzee short-length reverse | 5′-GAGTTGATGGGGTGCTCAG-3′ | Exon 3–intron 3 junction | |
| Chimpanzee short-length probe | 5′-GAGTTGATGGGGTGCTCAG-3′ | Exon 3 | |
Four more transcripts were identified in the chimpanzee. Chimpanzee PSA-3 has an open reading frame of 456 bases. It skips 89 bases of exon 3, resulting in an early stop codon. It codes for a 151 amino acid protein and has a predicted molecular weight of 16,470 Da. It is novel, as no similar transcript was found in the NCBI GenBank.

Chimpanzee PSA-4 is a result of an in-frame deletion of 129 bases of exon 3. This transcript has an open reading frame of 657 bases, which encodes a 218 amino acid protein, and has a molecular weight of 23,768 Da. This transcript is orthologous to the human \textit{KLK3} transcript variant 4 (NCBI accession number NM_001030048) and a rhesus monkey sequence (UniProt number F6TBJ4_MACMU).

Chimpanzee PSA-5 is novel, consisting of four exons. It skips 36 bases in the 5' part of exon 4, resulting in an early stop codon, and is therefore shorter than chimpanzee PSA-1. It has an open reading frame of 507 bases and encodes a 168 amino acid protein and has a predicted molecular weight of 18,774 Da.

Chimpanzee PSA-6 is also novel. It is similar to the chimpanzee PSA-2 transcript, except that it skips 129 bases in the 5' part of exon 3 as a result of use of an alternative splice acceptor site. It has an open reading frame of 414 bases, which encodes a 137 amino acid protein, and has a molecular weight of 14,730 Da.

We cannot rule out the possibility that other splice variants were not found by our cloning strategy, especially those expressed at much lower levels.

All the novel transcripts reported in this study for nonhuman primates were also found in human prostate cDNA, indicating greater similarity among the prostate transcriptomes across primates than has previously been appreciated. In addition, all the \textit{PSA} transcripts reported have an intact signal peptide consisting of 17 amino acids. However, only the PSA-1 transcripts have an intact catalytic triad.

A comparison of the amino acid sequences of the PSA-1 and PSA-2 transcripts is shown in Figure 2. The transcripts reported in this study have been deposited in the NCBI GenBank under accession numbers KC853005 (\textit{Chlorocebus aethiops} PSA-1),

\begin{table}
\centering
\begin{tabular}{|l|l|}
\hline
\textbf{PSA Variant} & \textbf{Splicing Event} \\
\hline
Chimpanzee (6 transcripts) & \\
Chimpanzee PSA-1 & Complete 5 exons \\
Chimpanzee PSA-2 & Uses sequences in intron 3 \\
Chimpanzee PSA-3 & Skips 89 bases in exon 3 \\
Chimpanzee PSA-4 & Skips 129 bases in exon 3 \\
Chimpanzee PSA-5 & Skips 36 bases in exon 4 \\
Chimpanzee PSA-6 & Skips 129 bases in exon 3 and uses sequences in intron 3 \\
\hline
Cynomolgus monkey (2 transcripts) & \\
Cynomolgus PSA-1 & Complete 5 exons \\
Cynomolgus PSA-2 & Uses sequences in intron 3 \\
\hline
Baboon (2 transcripts) & \\
Baboon PSA-1 & Complete 5 exons \\
Baboon PSA-2 & Uses sequences in intron 3 \\
\hline
African green monkey (2 transcripts) & \\
African green monkey PSA-1 & Complete 5 exons \\
African green monkey PSA-2 & Uses sequences in intron 3 \\
\hline
\end{tabular}
\end{table}
We next investigated quantitative differences in expression levels among the six chimpanzee transcripts. Using real-time PCR techniques, we did not find a significant difference in amplification rates between the transcripts that use sequences in intron 3 (chimpanzee PSA-2 and PSA-6) from those that do not (chimpanzee PSA-1, -3, -4, -5) (Figure 3).

Phylogenetic Analysis

We compared the PSA gene sequences among 11 primate species to analyze selection on the PSA gene. The pairwise dN/dS values across all codon sites of the PSA gene of any two species did not exceed the threshold of 1, indicating purifying selection (Table S1). Although purifying selection dominated across all comparisons, multiple regions in the C-terminal region appeared to
under relaxed or positive selection with dN/dS (v) values greater than 2. To identify individual residues under positive selection, we used the PAML software package, version 4.7, with the following comparisons: M0 vs. M3, M1 vs. M2, and M7 vs. M8 using the BEB and NEB approaches and the likelihood ratio test. All model comparisons with CODEML identified multiple residues with \( p > 0.5 \) (probability of selection greater than 0.5), indicating directional selection. The M7 vs. M8 comparisons using the NEB and BEB approaches were similar, identifying five codons with dN/dS values greater than 2 (Table 4; Figure 4). When the codon-specific dN/dS value data were compared with the alternative splicing data, we found that several splice variants originated at codons with dN/dS values greater than 2 (Table 4).

All methods generated similar consensus trees with similar bootstrap support for internal nodes. The only exception was the branching of *Gorilla gorilla* and *Nomascus leucogenys*, which varied between methods. The maximum-likelihood phylogenetic tree is shown in Figure 5, and the branching order is similar to the topology reported by others, except for the placement of *Gorilla gorilla* [20]. We looked at branch-specific variation in selective pressure along phylogenetic lineages. This did not reveal any statistically significant differences, but we noted that the variations in dN/dS values ranged from 0 to 0.76 (Figure 5).

### Discussion

We found six alternatively spliced *PSA* variants in the chimpanzee but only two in the cynomolgus monkey, baboon, and African green monkey. To the best of our knowledge, this study is the first to report the African green monkey *PSA* mRNA sequence and also the first to report on alternative splicing of the *PSA* gene in nonhuman primates. It is important to note that the identification of alternative splice variants depends on the primers used, and therefore we cannot rule out the possibility that there are other splice variants not found by our cloning primers.

Alternative splicing of the *PSA* gene in this study involved exon truncation, intron retention, or a combination of two splicing events. In a study of human kallikrein genes, Kurlender et al. (2005) found exon skipping to be the most common splicing event, with internal exon deletion being less common in kallikrein genes [34]. However, in our study we found retention of sequences in intron 3 to be the most common splicing event; it was found in all the species studied. Real-time PCR data indicated that in chimpanzees the expression of transcripts that do not use sequences in intron 3 did not take precedence over those that do, although we cannot rule out the possibility that under certain physiological conditions there might be differences in transcript expression. It is not clear why alternative splicing of the *PSA* gene is more common in chimpanzees than in cynomolgus monkeys, baboons, and African green monkeys, but chimpanzees do share a more recent common ancestor with humans than any other nonhuman primates do. All the transcripts reported here have orthologous human counterparts. The *PSA-2* sequences are orthologous to the human *PSA-RP2* sequence, which has been shown to be up-regulated in prostate cancer [35], indicating potential links of these transcripts to prostatic diseases. The link between these transcripts and human disease conditions needs further study. It is important to note that the splice variants were found in areas of the gene that are likely to be under selective pressure.

Not surprisingly, given the role of PSA, purifying selection dominates in all pairwise comparisons. Further analysis using codon-based models in CODEML identified multiple residues in the *PSA* gene having dN/dS values greater than 1, indicating directional selection. These data suggest that selective pressure in the *PSA* gene is not uniform across all codon sites but rather is focused on specific regions. Five codons were found to have dN/dS values greater than 2, although only one His45 reached the 95% confidence level. It is important to note that another study also found codons 45, 189, and 203 to be positively selected [20]. In addition, we note that limiting our analysis to exons 1–4 curtailed our statistical power. When the same analyses were

### Table 4. Bayes empirical Bayes likelihood ratio test of *PSA* gene codons that are under positive selection using model M7 vs. model M8.

| Position and Residue Change (KLK3 Numbering) | Position and Residue Change (Chymotrypsin Numbering) | dN/dS | Probability of dN/dS > 1 | Splice Variant Generated at Site |
|---------------------------------------------|------------------------------------------------------|-------|--------------------------|---------------------------------|
| H45R                                        | H37R                                                 | 4.40* | 0.98                     | Chimpanzee PSA-6, Chimpanzee PSA-4 |
| N69K/S                                      | N61K/S                                                | 2.64  | 0.6                      |                                 |
| Q144W                                       | Q128W                                                 | 2.60  | 0.60                     |                                 |
| P189S                                       | P173S                                                 | 4.30  | 0.94                     |                                 |
| T203M/R                                     | T187M/R                                               | 3.51  | 0.77                     |                                 |

* p < 0.05.

doi:10.1371/journal.pone.0094522.t004
performed using full-length sequences with all five exons, we found additional codons that had $dN/dS$ values greater than 1. Limiting our analyses to the first four exons was due to the reported fusion of $KLK3$ and $KLK2$ in $Gorilla gorilla$ and $Nomascus leucogenys$, in which the first four exons of the fused gene are orthologous to $KLK3$ and the last exon is more similar to $KLK2$ [20].

At codon position 189 the amino acid proline is substituted for serine or phenylalanine. Substitution of proline residues in proteins is of significant interest because of proline’s unique structural and functional properties. Proline has a rigid conformation with low flexibility; it can often be found in tight turns in protein structures and also introduces kinks into alpha helices because it is unable to adopt a normal helical conformation. It is likely that proline substitution at codon 189 results in secondary and tertiary structural alteration, leading to significant functional changes. Although we did not explore functional consequences of the change at codon position 189 in this study, this is an excellent question for future analysis.

Figure 4. Multiple sequence comparison by log expectation (MUSCLE) program guided alignments of PSA-1 transcripts from 11 species. The codons that are under positive selection are indicated by arrows. Fully conserved residues are indicated by asterisks, partly conserved residues by colons, and weakly conserved residues by periods. The species’ complete scientific names are as follows: $Pan troglodytes$, $Pan paniscus$, $Homo sapiens$, $Pongo pygmaeus$, $Gorilla gorilla$, $Nomascus leucogenys$, $Macaca fascicularis$, $Macaca fuscata$, $Macaca mulatta$, $Papio hamadryas anubis$, and $Chlorocebus aethiops$.

doi:10.1371/journal.pone.0094522.g004

confirmed using full-length sequences with all five exons, we found additional codons that had $dN/dS$ values greater than 1. Limiting our analyses to the first four exons was due to the reported fusion of $KLK3$ and $KLK2$ in $Gorilla gorilla$ and $Nomascus leucogenys$, in which the first four exons of the fused gene are orthologous to $KLK3$ and the last exon is more similar to $KLK2$ [20].

At codon position 189 the amino acid proline is substituted for serine or phenylalanine. Substitution of proline residues in proteins is of significant interest because of proline’s unique structural and functional properties. Proline has a rigid conformation with low flexibility; it can often be found in tight turns in protein structures and also introduces kinks into alpha helices because it is unable to adopt a normal helical conformation. It is likely that proline substitution at codon 189 results in secondary and tertiary structural alteration, leading to significant functional changes. Although we did not explore functional consequences of the change at codon position 189 in this study, this is an excellent question for future analysis.

Figure 4. Multiple sequence comparison by log expectation (MUSCLE) program guided alignments of PSA-1 transcripts from 11 species. The codons that are under positive selection are indicated by arrows. Fully conserved residues are indicated by asterisks, partly conserved residues by colons, and weakly conserved residues by periods. The species’ complete scientific names are as follows: $Pan troglodytes$, $Pan paniscus$, $Homo sapiens$, $Pongo pygmaeus$, $Gorilla gorilla$, $Nomascus leucogenys$, $Macaca fascicularis$, $Macaca fuscata$, $Macaca mulatta$, $Papio hamadryas anubis$, and $Chlorocebus aethiops$.

doi:10.1371/journal.pone.0094522.g004

confirmed using full-length sequences with all five exons, we found additional codons that had $dN/dS$ values greater than 1. Limiting our analyses to the first four exons was due to the reported fusion of $KLK3$ and $KLK2$ in $Gorilla gorilla$ and $Nomascus leucogenys$, in which the first four exons of the fused gene are orthologous to $KLK3$ and the last exon is more similar to $KLK2$ [20].

At codon position 189 the amino acid proline is substituted for serine or phenylalanine. Substitution of proline residues in proteins is of significant interest because of proline’s unique structural and functional properties. Proline has a rigid conformation with low flexibility; it can often be found in tight turns in protein structures and also introduces kinks into alpha helices because it is unable to adopt a normal helical conformation. It is likely that proline substitution at codon 189 results in secondary and tertiary structural alteration, leading to significant functional changes. Although we did not explore functional consequences of the change at codon position 189 in this study, this is an excellent question for future analysis.

Figure 4. Multiple sequence comparison by log expectation (MUSCLE) program guided alignments of PSA-1 transcripts from 11 species. The codons that are under positive selection are indicated by arrows. Fully conserved residues are indicated by asterisks, partly conserved residues by colons, and weakly conserved residues by periods. The species’ complete scientific names are as follows: $Pan troglodytes$, $Pan paniscus$, $Homo sapiens$, $Pongo pygmaeus$, $Gorilla gorilla$, $Nomascus leucogenys$, $Macaca fascicularis$, $Macaca fuscata$, $Macaca mulatta$, $Papio hamadryas anubis$, and $Chlorocebus aethiops$.

doi:10.1371/journal.pone.0094522.g004
Our results of codon-based selection indicate that dN/dS values for the PSA gene are low relative to other proteins from the seminal fluid and prostasome. Clark and Swanson (2005) reported a dN/dS value of 2.9 for β-microseminoprotein and a value of 14 for prostate-specific transglutaminase for the class of codons identified to be under positive selection [21]. It is possible that the critical role for the PSA protein in copulatory plug dissolution constrains the changes tolerated and maintains low dN/dS values. Alternatively, the atypical evolutionary history of the PSA gene (e.g., fusion of KLK3 and KLK2 in some lineages) may be playing a contributing role that has yet to be defined.

Previous studies have indicated that chimpanzees have low serum PSA concentrations [36]. In our unpublished data we also found that chimpanzee PSA concentrations range from 0.01 to 0.031 ng/ml, well below levels reported in humans and other nonhuman primate species [22]. We propose that PSA gene evolution in the chimpanzee lineage might have functional implications. Clark and Swanson (2005) also showed that although the PSA gene does not have high pairwise human-chimpanzee dN/dS values, it does show significant variation in selective pressure during its evolution and dN/dS values along some phylogenetic lineages exceeded 1 [21].

The placement of the gorilla in our phylogenetic tree may be due to incomplete lineage sorting, whereby the genealogy relating humans, gorillas, and chimpanzees varies across the genome and about 22% of all coding exons exhibit phylogenies that contradict the overall genomic phylogeny, in which humans are more similar to chimpanzees than to gorillas [37]. The genomic region containing PSA may be one of the regions that has a local phylogeny in which the gorilla sequences are indeed more similar to the human sequences than the chimpanzee sequences are to either human or gorilla.

Because the function of PSA is to liquefy the sperm coagulum formed after ejaculation, we speculate that the differences in alternative splicing reported in this study might be related to the mating systems of these species. The four species studied here (chimpanzees, cynomolgus monkeys, baboons, and African green monkeys) all exhibit multimale-multifemale mating systems in which any given adult female may mate with more than one male during a single estrous cycle. However, some data suggest that chimpanzee females may have larger average numbers of mates per cycle. A single chimpanzee female frequently mates with numerous males during a single estrous cycle [38,39]. Multiple males may follow, compete for, and copulate with a female during a single cycle [39]. Such male-male competition also occurs in other species to some degree. But there is reason to believe that sperm competition is important in chimpanzees, as a rubbery and long-lasting “copulatory plug” that obstructs the sperm of subsequent mating events from accessing the ovum is produced by chimpanzee ejaculates and not by those of cynomolgus monkeys, African green monkeys, and baboons [40]. Further studies of transcript heterogeneity comparing several species that have different mating patterns would further elucidate the possible relationship between the alternative expression of the PSA gene and mating systems.

Supporting Information

Table S1 Species pairwise comparison of the dN/dS values of the PSA gene.

Acknowledgments

We would like to thank Dr. Matthew J. Jorgensen of the Department of Pathology/Comparative Medicine, Wake Forest University Health Sci-
References

1. Yousef GM, Chang A, Scorrillas A, Diamandis EP (2000) Genomic organization of the PSA gene family on chromosome 19q13.3-13.4. Biochem Biophys Res Commun 276: 125–133.

2. Diamandis EP, Yousef GM, Loo I, Maglaka I, Obici CV (2000) The new human kallikrein gene family: Implications in carcinogenesis. Tr Endocrinol Metab 11: 54–60.

3. Lundwall A, Lilja H (1987) Molecular cloning of human prostate specific antigen cDNA. FEBS Lett 214: 317–322.

4. Riegman PH, Vliestru RJ, van der Korput JA, Trapman J (1989) Characterization of the prostate-specific antigen gene: A novel human kallikrein-like gene. Biochem Biophys Res Commun 159: 95–102.

5. Carter P, Wells J (1980) Dissecting the catalytic triad of a serine protease. Nature 332: 564–568.

6. Riegman PH, Klaassen P, van der Korput JA, Romijn JC, Trapman J (1988) Molecular cloning and characterization of novel prostate antigen cDNAs. Biochem Biophys Res Commun 155: 181–188.

7. Balk SP, Ko VJ, Bubley GJ (2003) Biology of prostate-specific antigen. J Clin Invest 80: 281–285.

8. Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc B 57: 289–300.

9. Yamada Y, Kato T, Yoshioka M, Hasegawa T, Tagawa T (1992) Expression of the hK2 gene in rat prostate. Biochem Biophys Res Commun 185: 986–990.

10. Sensabaugh GF (1978) Isolation and characterization of a semen-specific protein with prostatic specificity. Proc Nat Acad Sci U S A 75: 4745–4750.

11. Lilja H, Oldbring J, Rannevik G, Laurell CB (1987) Seminal vesicle–secreted seminal vesicle protein. J Clin Invest 76: 1899–1903.

12. Dorus S, Evans PD, Wyckoff GJ, Choi SS, Lahn BT (2004) Rate of molecular evolution in the fast lane: Rapidly evolving sex-related genes in Drosophila. Trends Genet 20: 201–205.

13. Sensabaugh GF (1978) Isolation and characterization of a semen-specific protein with prostatic specificity. Proc Nat Acad Sci U S A 75: 4745–4750.

14. Lilja H (1985) A kallikrein-like serine protease in prostatic fluid cleaves the predominant seminal plasma proteins and their reactions during gelation and liquefaction of human semen. J Clin Invest 76: 1899–1903.

15. Wang ET, Sandberg R, Luo S, Khrebtukova I, Zhang L, et al. (2008) Evolutionary trees from DNA sequences: A maximum likelihood approach. J Mol Evol 17: 568–576.

16. Felsenstein J (1981) Evolutionary trees from DNA sequences: A maximum likelihood approach. J Mol Evol 17: 568–576.

17. Heuze-Vourc’h N, Leblond V, Olayat S, Gauthier F, Courty Y (2001) Complex alternative splicing of the PSA gene coding for the tumor marker PSA (prostate-specific-antigen). Eur J Biochem 268: 792–800.

18. Heuze-Vourc’h N, Leblond V, Olayat S, Gauthier F, Courty Y (2001) Complex alternative splicing of the PSA gene coding for the tumor marker PSA (prostate-specific-antigen). Eur J Biochem 268: 792–800.

19. Kurlender L, Borgono C, Michael IP, Obiezu C, Elliott MB, et al. (2005) Specificity profiling of seven human tissue kallikreins reveals individual subsite preferences. J Biol Chem 280: 25670–25680.

20. Marshall DJ, Rudnick KA, Lu J, Snyder LA (2006) Cloning and sequencing of the cynomolgus monkey prostate specific antigen cDNA. J Med Primatol 35: 12–17.

21. Clark NL, Swanson WJ (2005) Pervasive adaptive evolution in primate seminal proteins. PLoS Genet 1:e35.

22. Mubiru JN, Hubbard GB, Dick EJ Jr, Furman J, Troyer DA, et al. (2000)Nonhuman primates as models for studies of prostate specific antigen and prostatic diseases. Prostate 68: 1546–1544.

23. Mubiru JN, Shen-ong GL, Valente AJ, Troyer DA (2004) Alternative spliced variants of the alpha-methylacyl-CoA racemase gene and their expression in prostate cancer. Gene 327: 89–98.

24. Steinway SN, Dannenfelser R, Laucius CD, Hayes JE, Nayak S (2010) JCoDA: A tool for detecting evolutionary selection. BMC Bioinformatics 11: 284.

25. Yang Z (2007) PAML 4: Phylogenetic analysis by maximum likelihood. Mol Biol Evol 18: 1588–1591.

26. Goldman N, Yang Z (1994) A codon-based model of nucleotide substitution for protein-coding DNA sequences. Mol Biol Evol 11: 725–736.

27. Yang Z, Bielawski JP (2000) Statistical methods for detecting molecular adaptation. Tr Evol 15: 496–509.

28. Yang Z, Wong WS, Nielsen R (2005) Bayes empirical Bayes inference of amino acid sites under positive selection. Mol Biol Evol 22: 1107–1118.

29. Goldman N, Yang Z (1994) A codon-based model of nucleotide substitution for protein-coding DNA sequences. Mol Biol Evol 11: 725–736.

30. Yang Z (2007) PAML 4: Phylogenetic analysis by maximum likelihood. Mol Biol Evol 18: 1588–1591.

31. Yang Z, Wong WS, Nielsen R (2005) Bayes empirical Bayes inference of amino acid sites under positive selection. Mol Biol Evol 22: 1107–1118.

32. Felsenstein J (1981) Evolutionary trees from DNA sequences: A maximum likelihood approach. J Mol Evol 17: 568–576.

33. Debela M, Magloden V, Schechter N, Valachova M, Luttospeich F, et al. (2006) Specificity profiling of seven human tissue kallikreins reveals individual subsite preferences. J Biol Chem 281: 25670–25680.

34. Kurlender L, Borgono C, Michael IP, Obiezu C, Elliot MB, et al. (2005) A survey of alternative transcripts of human tissue kallikrein genes. Biochem Biophys Acta 25: 1–14.

35. Whitbread AK, Vererose TL, Dong Y, Tan OL, Gardiner R, et al. (2010) Expression of PSA-RP1, an alternatively spliced variant from the PSA gene, is increased in prostate cancer tissues but the protein is not secreted from prostate cancer cells. Biol Chem 391: 461–466.

36. Steiner MS, Couch RC, Raghoe S, Stauffer D (1999) The chimpazee as a model of human benign prostatic hyperplasia. J Urol 162: 1454–1461.

37. Scally A, Dutheil JY, Hillier LW, Jordan GE, Goodhead I, et al. (2012) Insights into hominid evolution from the gorilla genome sequence. Nature 483: 169–175.

38. Dixon AF (2012) Primate sexuality: Comparative studies of the prosimians, monkeys, apes, and humans. New York: Oxford University Press.

39. Goodall J (1986) The chimpanzees of Gombe: Patterns of behavior. Cambridge, MA: Belknap Press, Harvard University Press.

40. Dixon AF, Anderson MJ (2002) Sexual selection, seminal coagulation, and copulatory plug formation in primates. Folia Primatol (Basel) 73: 63–69.

41. Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTALW: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res 11: 4673–4680.

Author Contributions

Conceived and designed the experiments: JNM JR RES EJD MO. Performed the experiments: JNM ASY CBL MGF MO EJD. Analyzed the data: JNM CO CBL JR RES SN. Contributed reagents/materials/analysis tools: CO CBL JR RES SN. Wrote the paper: JNM JR CO SN. Obtained the software to run analysis: SN CBL CO.