MIT Open Access Articles

Sequencing of rhesus macaque Y chromosome clarifies origins and evolution of the DAZ (Deleted in AZoospermia) genes

The MIT Faculty has made this article openly available. ***Please share*** how this access benefits you. Your story matters.

Citation	Hughes, Jennifer F., Helen Skaletsky, and David C. Page. "Sequencing of rhesus macaque Y chromosome clarifies origins and evolution of the DAZ (Deleted in AZoospermia) genes." BioEssays 34, no. 12 (December 10, 2012): 1035-1044.
As Published	http://dx.doi.org/10.1002/bies.201200066
Publisher	John Wiley & Sons, Inc
Version	Author's final manuscript
Citable link	http://hdl.handle.net/1721.1/85095
Terms of Use	Creative Commons Attribution-Noncommercial-Share Alike
Detailed Terms	http://creativecommons.org/licenses/by-nc-sa/4.0/

Massachusetts Institute of Technology

DSpace@MIT

NIH Public Access
Author Manuscript
Bioessays. Author manuscript; available in PMC 2013 December 01.

Published in final edited form as:
Bioessays. 2012 December ; 34(12): 1035–1044. doi:10.1002/bies.201200066.

Sequencing of rhesus macaque Y chromosome clarifies origins and evolution of the *DAZ* (*Deleted in AZoospermia*) genes

Jennifer F. Hughes, **Helen Skaletsky**, and **David C. Page**
Howard Hughes Medical Institute, Whitehead Institute, and Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02142 jhughes@wi.mit.edu

Abstract

Studies of Y chromosome evolution often emphasize gene loss, but this loss has been counterbalanced by addition of new genes. The *DAZ* genes, which are critical to human spermatogenesis, were acquired by the Y chromosome in the ancestor of Old World monkeys and apes. We and our colleagues recently sequenced the rhesus macaque Y chromosome, and comparison of this sequence to human and chimpanzee enables us to reconstruct much of the evolutionary history of *DAZ*. We report that *DAZ* arrived on the Y chromosome about 36 million years ago via the transposition of at least 1.1 megabases of autosomal DNA. This transposition also brought five additional genes to the Y chromosome, but all five genes were subsequently lost through mutation or deletion. As the only surviving gene, *DAZ* experienced extensive restructuring, including intragenic amplification and gene duplication, and has been the target of positive selection in the chimpanzee lineage.

Keywords

Y chromosome; *DAZ*; rhesus macaque; chimpanzee; ampliconic

Introduction

X and Y chromosomes have evolved from autosomes repeatedly in diverse animal and plant species. A seemingly inevitable feature of this process is the degeneration and decay of the Y chromosome. This is because the Y chromosome has suppressed or reduced recombination with its meiotic partner, the X chromosome, and consequently has relinquished the evolutionary benefits of sexual recombination [1]. Because the Y chromosome is restricted to males, it is an attractive place for male-benefit genes to reside. Suppression of recombination is thus selectively advantageous because it ensures that these genes are always expressed in males but not females, where they may have detrimental effects [2]. However, this widespread suppression of recombination comes at a price: rapid gene loss and extensive deletions on the Y chromosome. The modern-day human Y chromosome bears the marks of hundreds of millions of years of these evolutionary processes. The male-specific region of the human Y chromosome (or MSY, which excludes the pseudoautosomal regions) has retained only ~3% of the genes it once shared with the X chromosome. Recently, full-scale comparative MSY sequencing in two additional primates – the chimpanzee and rhesus macaque – has shown that MSY ancestral gene content has remained stable in the human lineage for at least 25 million years, meaning that the pace of decay has dramatically slowed or even come to a halt [3–5].

While MSY decay has been the subject of empirical and theoretical studies, other evolutionary processes have also shaped the human MSY: sequence addition and amplification. Nearly half of the euchromatic human MSY is comprised of “ampliconic”

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

sequence, which is populated with genes that are exclusively expressed in the testis [6]. While some of the MSY's ampliconic genes date back to the ancestral autosomes from which the X and Y evolved, others have arrived more recently via transpositions from the X chromosome and autosomes. A unifying feature of these recently added genes, in addition to their multicopy nature and testis-specific expression pattern, is their high degree of divergence from their autosomal or X-linked predecessors. This divergence is likely a consequence of new or specialized functional roles (e.g. in spermatogenesis) acquired by these genes after they took up residence on the MSY.

No gene exemplifies these evolutionary transitions better than *DAZ* (*Deleted in AZoospermia*), which arrived on the MSY as part of a transposition from chromosome 3 in the Old World monkey (OWM) – ape ancestor (together, the catarrhine primates) [7]. There are four highly similar copies of *DAZ* on both the human and chimpanzee reference MSYs [4,6,8], and several lines of evidence suggest that the *DAZ* genes play an important role in spermatogenesis (Box 1). Comparison of Y-linked *DAZ* to its autosomal homologue, *DAZL*, reveals that the *DAZ* genes experienced an extraordinary degree of structural change in the Y lineage. We and our colleagues recently determined the MSY sequence of an OWM, the rhesus macaque [5], and this sequence allows us to see that *DAZ* followed a distinct evolutionary trajectory in the rhesus lineage. In light of the new rhesus MSY sequence, we now revisit and clarify the evolutionary history of the *DAZ* genes. For each of three major processes in *DAZ*'s history – transposition, intragenic amplification, and gene duplication – we will first present what was known based on human and chimpanzee sequence and then introduce novel insights based on our comparative analyses with the more distantly related primate, the rhesus macaque.

Box 1: Evidence for *DAZ's* role in spermatogenesis

Recurrent deletions on the Y chromosome are the leading genetic cause of spermatogenic failure in men [9]. Several of these deletions, which are caused by ectopic recombination between ampliconic repeats, remove some or all copies of *DAZ* [10–13]. However, these deletions ablate other testis-specific genes and gene families as well, and no deletions or mutations have been identified that only affect *DAZ*. Given the intermingled arrangement of the *DAZ* genes with other gene families [12] and *DAZ*'s multicopy nature, it is difficult to imagine a human Y-chromosome deletion that would remove all four *DAZ* genes while leaving all other Y-chromosome gene undisturbed. Fortunately, with recent advances in transgenics in rhesus macaque [14,15], meaningful study of *DAZ* function may soon be possible. In rhesus, there are only two copies of *DAZ* [5], potentially making targeted mutagenesis feasible. Rhesus spermatogenesis is very similar to that in human [16], so phenotypic insight gained from such studies will be medically relevant.

Additional evidence for involvement of *DAZ* in spermatogenesis comes from functional studies of its autosomal progenitor, *Dazl*, in mouse. Adult *Dazl* knockout mice of both sexes are infertile and have severely impaired germ cell development [17]. Indeed, studies of *Dazl*-deficient mice reveal that the gene plays a central role in fetal germ cell development in both sexes [18]. Furthermore, autosomal homologs of *DAZ* are conserved throughout metazoa and have been shown to be critical for fertility in *Drosophila*, *C. elegans*, *Xenopus*, and zebrafish [19–22].

Remainders of a massive autosomal transposition on the rhesus Y chromosome

DAZ was the first MSY gene shown to have arisen via transposition of an autosomal gene to the MSY [7]. Direct evidence that *DAZ* had an autosomal progenitor (*DAZL*) came from analyses of human testis cDNA clones mapping to chromosome 3 whose sequences were homologous to, but clearly distinct from, Y-linked *DAZ* [7]. Southern blot analyses revealed

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

that Y-linked *DAZ* was present in catarrhine primates but absent in New World monkeys (NWM) and more distantly related mammals [10,23,24], indicating that the chromosome 3 transposition event that brought *DAZ* to the MSY occurred 25–42 million years ago [25]. Outside the coding sequence of the *DAZ* gene, however, very little chromosome-3-homologous sequence was found on either the human MSY (Fig. 1) or chimpanzee MSY (data not shown).

The chromosome 3 transposition event occurred in the common ancestor of human, chimpanzee, and rhesus, so we analyzed the rhesus MSY sequence for traces of this event. We found that the rhesus MSY contains chromosome-3-homologous sequences that span roughly 1.1 megabases (Mb) of chromosome 3 (Fig. 1). This region of chromosome 3 contains not only *DAZL*, but also five other coding genes, and we found pseudogene remnants of four of these genes on the rhesus MSY. These observations lead us to two conclusions about the introduction of *DAZ* to the MSY. First, the transposed segment was at least 1.1 Mb in size, and roughly 8% of the euchromatic portion of the present-day rhesus MSY is derived from the chromosome 3 transposition (compared to only ~1% in human). This massive segmental duplication is ~20 times the average size of catalogued transpositions in primate genomes [26]. Second, after arriving on the MSY, five of the six genes within the transposed region decayed or were deleted.

These observations enabled by analysis of the rhesus MSY sequence illustrate anew an oft-repeated theme in Y chromosome evolution: gene addition followed by decay. About 105 million years ago, in the ancestor of eutherian mammals, a 47-Mb autosomal segment was translocated from an autosome to the shared, pseudoautosomal region of the X and Y chromosomes [27,28]. Following suppression of X-Y crossing over, this resulted in addition of 162 new genes to the MSY, yet only 12 remain on the present-day human MSY [5]. These surviving genes evidently persisted because of the action of purifying selection [3,5,29]. Similar scenarios have played out multiple times in *Drosophila* as well, where several instances of recent sex chromosome-autosome fusion events have generated “neo-Y” chromosomes [30], enabling the opportunity to study the short-term effects of Y-chromosome linkage and the resulting loss of recombination. The best characterized neo-Y system is that of *Drosophila miranda*, where massive gene loss has occurred in the short timespan of one million years [31].

Intragenic amplification and diversification

While the chromosome 3 transposition event carried a total of six genes to the MSY, only *DAZ* survived, likely because it acquired a novel function that was selectively advantageous to males. A new role for *DAZ* is suggested by its significant structural reorganization compared to *DAZL*, resulting in an increase in both overall size of the transcription unit (from 18.7 kilobases [kb] to 84.0 kb) and total exon number (from 11 to 50) (Fig. 2A). *DAZ* experienced two independent internal amplifications, creating a 2.4-kb genomic repeat and a 10.8-kb genomic repeat [7,8]. The 2.4-kb unit was apparently amplified before the 10.8-kb unit, as shown by the fact that some copies of the 10.8-kb unit also contain a partially duplicated 2.4-kb unit (e.g. human *DAZ1* in Fig. 2A). The boundaries of both repeat units are identical in human, chimpanzee, and rhesus, indicating that both internal amplifications occurred initially in the catarrhine ancestor. The 10.8-kb repeat encompasses exons 2–6, which encode the RNA-recognition motif or RRM domain. In *DAZL*, this domain has been implicated in translational regulation [32,33]. The 2.4-kb repeat encompasses exons 7 and 8, but only exon 7 contributes to the coding sequence. The encoded 24-amino-acid repeat has no known function but is extraordinarily polymorphic in copy number and sequence among men [34–36].

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

In human, *DAZ* also differs significantly from *DAZL* at the level of mRNA processing as a number of exons are selectively pruned from the *DAZ* transcript [7,8]. Exon 9 of *DAZ* is present in the genomic MSY sequence but excluded from the processed mRNA (Fig. 2A), even though its donor and acceptor splice sites are identical to those found in *DAZL*, where it is included in the processed mRNA. In addition, while both exons 7 and 8 are contained within the 2.4-kb repeat unit, only exon 7 is spliced into the *DAZ* transcript (Fig. 2A). Most of the exon 8 copies have retained the splice site dinucleotides present in *DAZL* (Fig. 2A), so the regulation of mRNA processing likely relies on more complex cues.

In rhesus macaque, there are two copies of *DAZ* with divergent mRNA processing patterns, and our comparison of these two sequences sheds light on the selective pruning process. One copy of rhesus *DAZ* displays the same splicing pattern seen in human, where the amplified copies of exon 8 are excluded from the processed mRNA (Fig. 2A). The second copy of rhesus *DAZ* incorporates the amplified copies of exon 8 into the processed mRNA, which lengthens the predicted protein by 136 amino acids. This second version of *DAZ* in rhesus may represent an intermediate state, having experienced the internal amplifications but not the deactivation of exon 8 within the repeats.

A comparison of these two versions of *DAZ* in rhesus helps us discern the more subtle sequence changes that are responsible for this differential splicing pattern. A variety of short exonic sequence motifs are known to either enhance or suppress splicing [37,38], but our analysis of the rhesus *DAZ* copies showed no correlation between the relative distribution of these motifs in exon 8 copies with inclusion or exclusion in mRNA processing. However, we did find a significant difference between included and excluded exon 8 copies when we compared the strengths of their 5′ splice sites, which encompass nine nucleotides at the exon-intron boundary, using maximum entropy analysis (Fig. 2B; P=0.0007, one-tailed Mann Whitney test) [39].

In rhesus, the exon 8-inclusive *DAZ* copy is transcribed more robustly than the selectively spliced copy [5], which might explain why a previous study found only the exon-8-containing *DAZ* in another OWM - *Macaca fascicularis*, or the cynomolgus macaque [40]. We have confirmed the presence of a second divergent copy of *DAZ* in this species using RT-PCR and sequencing of testis cDNA (Fig. 3A; GenBank accession number JQ975875). In addition, BLAST searches of male genomic DNA sequence from a more distantly related OWM (*Chlorocebus aethiops*, vervet monkey) revealed the presence of the same two divergent *DAZ* copies (Fig. 3A), indicating that the alternative patterns of *DAZ* transcription in rhesus are likely conserved in OWMs.

Gene duplication: homogenization vs. divergence

DAZ experienced not only extensive internal amplification but also whole-gene duplication. In both human and chimpanzee, there are two pairs of *DAZ* genes, or a total of four copies. Each *DAZ* pair is situated within a palindrome [4,6,8], which is a structure composed of two large mirror-image repeats, or arms, separated by a short non-repeat sequence. The palindromes containing *DAZ* in human and chimpanzee are very similar, but are not found in gorilla [41], so the initial formation of this palindrome might be a relatively recent event, occurring between six and nine million years ago [25]. The paired arms of MSY palindromes undergo frequent gene conversion, which maintains a strikingly high degree of sequence identity [6,41]. Within the *DAZ* palindromes, the arm-to-arm identity is 99.97% in human and 99.98% in chimpanzee [4,6]. Therefore, the *DAZ* copies within each species, despite exhibiting a high degree of structural variation (Fig. 2), are remarkably similar to each other at the nucleotide level (Figs. 4,5).

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

There are two copies of the *DAZ* palindrome on the reference MSYs of both human and chimpanzee. However, we conclude that the palindrome duplication events, which increased the *DAZ* copy number from two to four, occurred independently in each lineage. The two palindromes are essentially adjacent in human, but are located on opposite arms of the chromosome, and over 9 Mb apart, in chimpanzee. Furthermore, close inspection of this region in human revealed that the palindrome duplication event was caused by ectopic recombination between human-specific Alu elements [12], and comparative sequencing in chimpanzee [4] confirms the absence of the corresponding Alus. The parallel duplication of the *DAZ* palindrome in human and chimpanzee suggests strong selective pressure to optimize *DAZ* copy number and thereby maximize reproductive fitness [42]. This speculation is consistent with a recent study identifying a significant association between reduced *DAZ* copy number and clinically low sperm count in men [43].

We now conclude that *DAZ* was duplicated in the OWM ancestor as well. Some previous studies indicated that *DAZ* is single copy in OWMs [24,40], while another study found two distinct copies of *DAZ* in OWMs [44]. The complete MSY sequence confirms that there are precisely two copies of *DAZ* in rhesus, and we also found evidence for at least two *DAZ* copies in two other OWM species, the cynomolgus macaque and the vervet monkey. The genomic sequence of the rhesus MSY reveals that the *DAZ* gene is not contained within a palindrome or other type of ampliconic structure, as it is in human and chimpanzee. Phylogenetic analyses indicate that the two copies of *DAZ* in rhesus do not experience extensive sequence homogenization and have diverged from each other (Fig. 3). In a phylogenetic tree generated using *DAZ* mRNA sequences, the two copies of *DAZ* in rhesus macaque, cynomolgus macaque, and vervet monkey form distinct and distant clusters (Fig. 3A), confirming that the two *DAZ* copies have been evolving independently in all three species. By contrast, the copies of *DAZ* in human and chimpanzee cluster in a species-dependent manner (Fig. 3). Since *DAZ* was duplicated prior to the human-chimpanzee split, this tree topology indicates ongoing gene conversion among all *DAZ* copies within the human MSY and the same within the chimpanzee MSY.

In rhesus, there is little evidence for gene conversion in the genomic sequence spanning the coding regions of the two *DAZ* genes, but gene conversion has still had an impact on the evolution of *DAZ*. The regions immediately upstream of each rhesus *DAZ* gene (and including the first coding exon) share 99.8% identity across 30 kb, which is markedly higher than the average identity of 86.2% found over the remainder of the genomic locus (Fig. 4A,D) and on par with the level of identity seen between *DAZ* copies in human (Fig. 4C). This striking degree of conservation may point to the presence of important regulatory signals within this region. The majority of this region is not conserved in human (Fig. 4B), implying divergence of transcriptional control of the *DAZ* family in human and rhesus. Both copies of rhesus *DAZ* display testis-specific expression [5], as in human. However, detailed studies of the developmental timing and cell-type-specific pattern of *DAZ* expression have yet to be conducted in rhesus, so more subtle species differences may indeed exist.

Evidence for positive selection in the chimpanzee lineage

A previous phylogenetic analysis of the *DAZ* genes in primates found evidence of purifying (negative) selection acting to conserve the encoded amino acid sequence across most of the gene [45]. We decided to specifically examine the RRM domain of *DAZ* for evidence of positive selection. Because the RRM domain is likely to serve a critical function, based on its amplification within the *DAZ* gene as well as experimental studies of the DAZL protein, this domain is a possible target for positive selection to alter the protein′s function. Therefore, we performed a more detailed phylogenetic study of this region of *DAZ* to look for evidence of positive selection, namely an elevated rate of nonsynonymous substitutions (dn) compared to synonymous substitutions (ds). Each of the *DAZ* genes in human,

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

chimpanzee, and rhesus contains between one and five copies of the RRM domain encoded by exons 2–6 (Fig. 2A and data not shown). We analyzed all copies of the coding sequences spanning the RRM repeat in each of the three species (20 sequences in total) using human *DAZL* sequence as an outgroup (Fig. 3C).

The results of this analysis provide further insights into both the dynamics of gene conversion within the *DAZ* genes and the impact of selection. The resulting tree topology shows tight species-specific clustering of the RRM repeats in human and chimpanzee (Fig. 3C), indicating that, in these species, intragenic (between-repeat) gene conversion may be acting in concert with intergenic gene conversion to maintain a high degree of identity between all RRM repeats within the *DAZ* genes. By contrast, the RRM repeats from the two *DAZ* genes in rhesus form separate clusters, and the repeats in rhesus *DAZ1* have diverged substantially from each other compared to the other gene copies. We then tested for evidence of positive selection in each branch of the phylogenetic tree [46], and found that the chimpanzee lineage shows a statistically significant elevation of the dn/ds ratio (2.709, $p = 0.004$; likelihood ratio test [47]). We previously speculated that positive selection targeting spermatogenesis factors on the Y chromosome has been more intense in the chimpanzee lineage than in human or rhesus [3–5]. Genetic hitchhiking associated with such positive selection provides a possible explanation for the acceleration of gene loss observed in the chimpanzee lineage [2]. Our new analysis indicates that the putative RNA-binding domain of the DAZ proteins has been the target of positive selection in the chimpanzee lineage.

A more complete history

With three MSY sequences in hand - human, chimpanzee, and rhesus macaque - we now have a 25-million-year view into the history of *DAZ* on the MSY and the ability to construct a more detailed model of the major events that shaped these genes (Fig. 5). It was previously recognized that *DAZ* transposed to the MSY from chromosome 3 in the catarrhine primate lineage after the divergence of catarrhines from NWMs but before the divergence of OWMs and apes [10,23] (between 25 and 42 million years ago [25]). We now know that this transposition event was at least 1.1 Mb in size and that *DAZ* was accompanied by five other chromosome-3-homologous genes, all of which have since decayed or been deleted. Only *DAZ*, which acquired an important male-specific function, survived. Molecular clock dating of the transposition event is now possible using the sizeable chromosome-3 homologous sequence found on the rhesus MSY. The degree of nucleotide divergence within neutrally-evolving intragenic sequence between the MSY and chromosome 3 correlates with the time since transposition, yielding an estimate of ~38.5 million years ago (Supplementary Note). This estimate places the timing of the transposition event shortly after the NWM-catharrine split, validating and refining previous estimates based on genomic hybridization [10,23]. Molecular clock analyses based on human or chimpanzee sequence would not be as accurate because the high degree of gene conversion experienced by *DAZ* in these species might reduce the rate of molecular evolution [41] (Fig. 3) and bias the estimate.

It is also now evident that *DAZ* was first duplicated relatively early in its evolutionary history – before the OWM-ape split (Fig. 5). We conclude that this event occurred in the common ancestor based on the fact that rhesus has retained two divergent *DAZ* genes. One copy is clearly more ancestral because it retains *DAZL*-like splicing of exon 8 within the 2.4-kb repeat. The more derived copy, which lost splicing of the amplified exon 8, is the only version of the gene that survived in the human/chimpanzee lineage, where it was subsequently amplified. Unless *DAZ* lost exon 8 splicing independently in the OWM and ape lineages, which seems improbable, we can conclude that *DAZ* was first duplicated in the catarrhine ancestor. Using molecular clock estimates based on the divergence of the two

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

DAZ copies in rhesus, we estimate that this initial duplication event occurred ~33 million years ago (Supplementary Note).

Conclusions

Assembling the MSY sequence of an OWM, the rhesus macaque [5], has contributed substantially to our view of the evolution of the catarrhine Y chromosome, and in particular, the *DAZ* genes. Because we completed the rhesus MSY sequence to the same rigorous standards as the human and chimpanzee MSYs [4,6], we were able to reconstruct and date important evolutionary events and processes that shaped this multicopy gene family. Since the rhesus macaque is among the best available model systems in which to study the function of *DAZ*, knowledge of the rhesus *DAZ* sequence will be a valuable resource to reproductive biologists. Human, chimpanzee, and rhesus represent only three lineages of over 150 catarrhine primate species. Yet, even among these three species, we see tremendous variation in the evolutionary trajectory of the *DAZ* genes, with possible implications for transcriptional regulation and protein function. Comparative sequencing in additional branches of the catarrhine tree will further expand our understanding of this biologically important gene family and its fascinating evolutionary history.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank W. Warren and K. Dewar at the Genome Institute at Washington University for providing genomic sequence from a male vervet monkey. We thank W. Bellott, C. Burge, R. Desgraz, G. Dokshin, Y.C. Hu, M. Kojima, B. Lesch, and K. Romer for useful discussion and comments on the manuscript. This work was supported by the National Institutes of Health, the Howard Hughes Medical Institute, and the Charles A. King Trust.

LITERATURE CITED

1. Barton NH, Charlesworth B. Why sex and recombination? Science. 1998; 281:1986–90. [PubMed: 9748151]
2. Rice WR. Genetic hitchhiking and the evolution of reduced genetic activity of the Y sex chromosome. Genetics. 1987; 116:161–7. [PubMed: 3596229]
3. Hughes JF, Skaletsky H, Pyntikova T, Minx PJ, Graves T, et al. Conservation of Y-linked genes during human evolution revealed by comparative sequencing in chimpanzee. Nature. 2005; 437:100–3. [PubMed: 16136134]
4. Hughes JF, Skaletsky H, Pyntikova T, Graves TA, van Daalen SK, et al. Chimpanzee and human Y chromosomes are remarkably divergent in structure and gene content. Nature. 2010; 463:536–9. [PubMed: 20072128]
5. Hughes JF, Skaletsky H, Brown LG, Pyntikova T, Graves T, et al. Strict evolutionary conservation followed rapid gene loss on human and rhesus Y chromosomes. Nature. 2012; 482:82–6. [PubMed: 22367542]
6. Skaletsky H, Kuroda-Kawaguchi T, Minx PJ, Cordum HS, Hillier L, et al. The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes. Nature. 2003; 423:825–37. [PubMed: 12815422]
7. Saxena R, Brown LG, Hawkins T, Alagappan RK, Skaletsky H, et al. The *DAZ* gene cluster on the human Y chromosome arose from an autosomal gene that was transposed, repeatedly amplified and pruned. Nat Genet. 1996; 14:292–9. [PubMed: 8896558]
8. Saxena R, de Vries JW, Repping S, Alagappan RK, Skaletsky H, et al. Four DAZ genes in two clusters found in the AZFc region of the human Y chromosome. Genomics. 2000; 67:256–67. [PubMed: 10936047]

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

9. Walsh TJ, Pera RR, Turek PJ. The genetics of male infertility. Semin Reprod Med. 2009; 27:124–36. [PubMed: 19247914]
10. Reijo R, Lee TY, Salo P, Alagappan R, Brown LG, et al. Diverse spermatogenic defects in humans caused by Y chromosome deletions encompassing a novel RNA-binding protein gene. Nat Genet. 1995; 10:383–93. [PubMed: 7670487]
11. Reijo R, J. S, Dinulos MB, Jaffe T, Brown LG, et al. Mouse autosomal homolog of *DAZ*, a candidate male sterility gene in humans, is expressed in male germ cells before and after puberty. Genomics. 1996; 35:346–52. [PubMed: 8661148]
12. Kuroda-Kawaguchi T, Skaletsky H, Brown LG, Minx PJ, Cordum HS, et al. The *AZFc* region of the Y chromosome features massive palindromes and uniform recurrent deletions in infertile men. Nat Genet. 2001; 29:279–86. [PubMed: 11687796]
13. Repping S, Skaletsky H, Lange J, Silber S, van der Veen F, et al. Recombination between palindromes P5 and P1 on the human Y chromosome causes massive deletions and spermatogenic failure. Am. J. Hum. Genet. 2002; 71:906–22. [PubMed: 12297986]
14. Niu Y, Yu Y, Bernat A, Yang S, He X, et al. Transgenic rhesus monkeys produced by gene transfer into early-cleavage-stage embryos using a simian immunodeficiency virus-based vector. Proc Natl Acad Sci U S A. 2010; 107:17663–7. [PubMed: 20870965]
15. Tachibana M, Sparman M, Ramsey C, Ma H, Lee HS, et al. Generation of chimeric rhesus monkeys. Cell. 2012; 148:285–95. [PubMed: 22225614]
16. Plant TM, Ramaswamy S, Simorangkir D, Marshall GR. Postnatal and pubertal development of the rhesus monkey (Macaca mulatta) testis. Ann N Y Acad Sci. 2005; 1061:149–62. [PubMed: 16467264]
17. Ruggiu M, Speed R, Taggart M, McKay SJ, Kilanowski F, et al. The mouse *Dazla* gene encodes a cytoplasmic protein essential for gametogenesis. Nature. 1997; 389:73–7. [PubMed: 9288969]
18. Gill ME, Hu YC, Lin Y, Page DC. Licensing of gametogenesis, dependent on RNA binding protein DAZL, as a gateway to sexual differentiation of fetal germ cells. Proc Natl Acad Sci U S A. 2011; 108:7443–8. [PubMed: 21504946]
19. Eberhart CG, Maines JZ, Wasserman SA. Meiotic cell cycle requirement for a fly homologue of human *Deleted in Azoospermia*. Nature. 1996; 381:783–5. [PubMed: 8657280]
20. Houston DW, King ML. A critical role for Xdazl, a germ plasm-localized RNA, in the differentiation of primordial germ cells in *Xenopus*. Development. 2000; 127:447–56. [PubMed: 10631166]
21. Karashima T, Sugimoto A, Yamamoto M. *Caenorhabditis elegans* homologue of the human azoospermia factor DAZ is required for oogenesis but not for spermatogenesis. Development. 2000; 127:1069–79. [PubMed: 10662646]
22. Hashimoto Y, Maegawa S, Nagai T, Yamaha E, Suzuki H, et al. Localized maternal factors are required for zebrafish germ cell formation. Developmental Biology. 2004; 268:152–61. [PubMed: 15031112]
23. Seboun E, Barbaux S, Bourgeron T, Nishi S, Algonik A, et al. Gene sequence, localization, and evolutionary conservation of DAZLA, a candidate male sterility gene. Genomics. 1997; 41:227–35. [PubMed: 9143498]
24. Yu YH, Lin YW, Yu JF, Schempp W, Yen PH. Evolution of the DAZ gene and the AZFc region on primate Y chromosomes. BMC Evol Biol. 2008; 8:96. [PubMed: 18366765]
25. Hedges SB, Dudley J, Kumar S. TimeTree: a public knowledge-base of divergence times among organisms. Bioinformatics. 2006; 22:2971–2. [PubMed: 17021158]
26. Cheng Z, Ventura M, She X, Khaitovich P, Graves T, et al. A genome-wide comparison of recent chimpanzee and human segmental duplications. Nature. 2005; 437:88–93. [PubMed: 16136132]
27. Ross MT, Grafham DV, Coffey AJ, Scherer S, McLay K, et al. The DNA sequence of the human X chromosome. Nature. 2005; 434:325–37. [PubMed: 15772651]
28. Waters PD, Duffy B, Frost CJ, Delbridge ML, Graves JA. The human Y chromosome derives largely from a single autosomal region added to the sex chromosomes 80–130 million years ago. Cytogenetics and Cell Genetics. 2001; 92:74–9. [PubMed: 11306800]

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

29. Rozen S, Marszalek JD, Alagappan RK, Skaletsky H, Page DC. Remarkably little variation in proteins encoded by the Y chromosome's single-copy genes, implying effective purifying selection. Am. J. Hum. Genet. 2009; 85:923–8. [PubMed: 20004767]
30. Charlesworth B, Charlesworth D. The degeneration of Y chromosomes. Philos. Trans. R. Soc. Lond. B Biol. Sci. 2000; 355:1563–72. [PubMed: 11127901]
31. Zhou Q, Bachtrog D. Sex-specific adaptation drives early sex chromosome evolution in Drosophila. Science. 2012; 337:341–5. [PubMed: 22822149]
32. Ruggiu M, Cooke HJ. In vivo and in vitro analysis of homodimerisation activity of the mouse Dazl1 protein. Gene. 2000; 252:119–26. [PubMed: 10903443]
33. Tsui S, Dai T, Warren ST, Salido EC, Yen PH. Association of the mouse infertility factor DAZL1 with actively translating polyribosomes. Biol Reprod. 2000; 62:1655–60. [PubMed: 10819768]
34. Ngo KY, Vergnaud G, Johnsson C, Lucotte G, Weissenbach J. A DNA Probe Detecting Multiple Haplotypes of the Human Y Chromosome. American Journal of Human Genetics. 1986; 38:407–18. [PubMed: 3010708]
35. Yen PH, Chai NN, Salido EC. The human *DAZ* genes, a putative male infertility factor on the Y chromosome, are highly polymorphic in the *DAZ* repeat regions. Mammalian Genome. 1997; 8:756–9. [PubMed: 9321470]
36. Jovelin F, Berthaud S, Lucotte G. Molecular basis of the TaqI p49a,f polymorphism in the DYS1 locus containing DAZ genes. Mol Hum Reprod. 2003; 9:509–16. [PubMed: 12900509]
37. Fairbrother WG, Yeh RF, Sharp PA, Burge CB. Predictive identification of exonic splicing enhancers in human genes. Science. 2002; 297:1007–13. [PubMed: 12114529]
38. Wang Z, Rolish ME, Yeo G, Tung V, Mawson M, et al. Systematic identification and analysis of exonic splicing silencers. Cell. 2004; 119:831–45. [PubMed: 15607979]
39. Yeo G, Burge CB. Maximum entropy modeling of short sequence motifs with applications to RNA splicing signals. J Comput Biol. 2004; 11:377–94. [PubMed: 15285897]
40. Gromoll J, Weinbauer GF, Skaletsky H, Schlatt S, Rocchietti-March M, et al. The Old World monkey *DAZ* (Deleted in AZoospermia) gene yields insights into the evolution of the *DAZ* gene cluster on the human Y chromosome. Hum Mol Genet. 1999; 8:2017–24. [PubMed: 10484770]
41. Rozen S, Skaletsky H, Marszalek JD, Minx PJ, Cordum HS, et al. Abundant gene conversion between arms of palindromes in human and ape Y chromosomes. Nature. 2003; 423:873–6. [PubMed: 12815433]
42. Repping S, van Daalen SK, Brown LG, Korver CM, Lange J, et al. High mutation rates have driven extensive structural polymorphism among human Y chromosomes. Nat Genet. 2006; 38:463–7. [PubMed: 16501575]
43. Noordam MJ, Westerveld GH, Hovingh SE, van Daalen SK, Korver CM, et al. Gene copy number reduction in the azoospermia factor c (AZFc) region and its effect on total motile sperm count. Hum Mol Genet. 2011; 20:2457–63. [PubMed: 21429917]
44. Agulnik AI, Zharkikh A, Boettger-Tong H, Bourgeron T, McElreavey K, et al. Evolution of the *DAZ* gene family suggests that Y-linked *DAZ* plays little, or a limited, role in spermatogenesis but underlines a recent African origin for human populations. Hum Mol Genet. 1998; 7:1371–7. [PubMed: 9700189]
45. Bielawski JP, Yang Z. Positive and negative selection in the DAZ gene family. Mol Biol Evol. 2001; 18:523–9. [PubMed: 11264403]
46. Yang Z. PAML: a program package for phylogenetic analysis by maximum likelihood. Computer Applications in the Biosciences. 1997; 13:555–6. [PubMed: 9367129]
47. Yang Z. Likelihood ratio tests for detecting positive selection and application to primate lysozyme evolution. Mol Biol Evol. 1998; 15:568–73. [PubMed: 9580986]
48. Felsenstein, J. PHYLIP (Phylogeny Inference Package) version 3.6. Distributed by the author. Department of Genome Sciences. University of Washington; Seattle: 2004. http://evolution.genetics.washington.edu/phylip.html
49. Mayor C, Brudno M, Schwartz JR, Poliakov A, Rubin EM, et al. VISTA : visualizing global DNA sequence alignments of arbitrary length. Bioinformatics. 2000; 16:1046–7. [PubMed: 11159318]

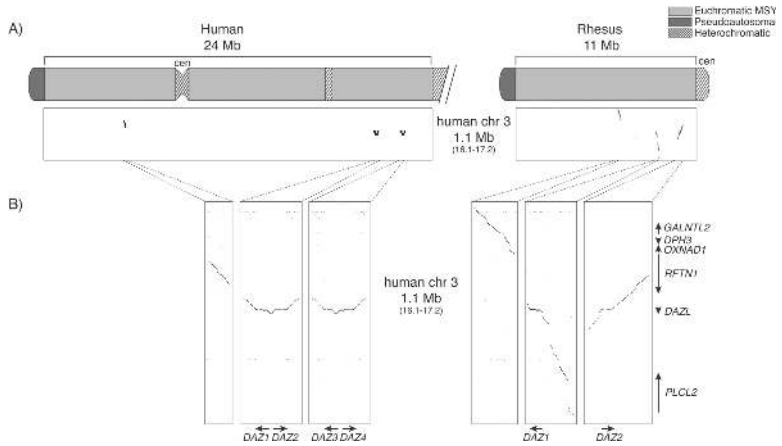

Figure 1.

Remnants of chromosome 3 transposition on rhesus and human MSYs. A: Dot-plot analyses of entire human (left) and rhesus (right) MSY sequences vs. 1.1-Mb region from human chromosome 3. Schematic representations of rhesus and human Y chromosomes are shown to scale. B: Separate dot-plot analyses of selected regions of human (left) and rhesus (right) MSY sequences vs. 1.1-Mb region from human chromosome 3. Position and orientation of Y chromosome *DAZ* genes and six chromosome 3 genes (including *DAZL*) are indicated by arrows on plot axes. For all plots, chromosome 3 sequence was masked with RepeatMasker (www.repeatmasker.org) prior to analysis. Each dot within the plot represents 100% identity within 30-bp window (A) or 20-bp window (B). Dot-plots were generated using custom perl code ("Fast Dot Plot" available at http://pagelab.wi.mit.edu/material-request.html). Abbreviations: MSY, male-specific Y; cen, centromere.

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

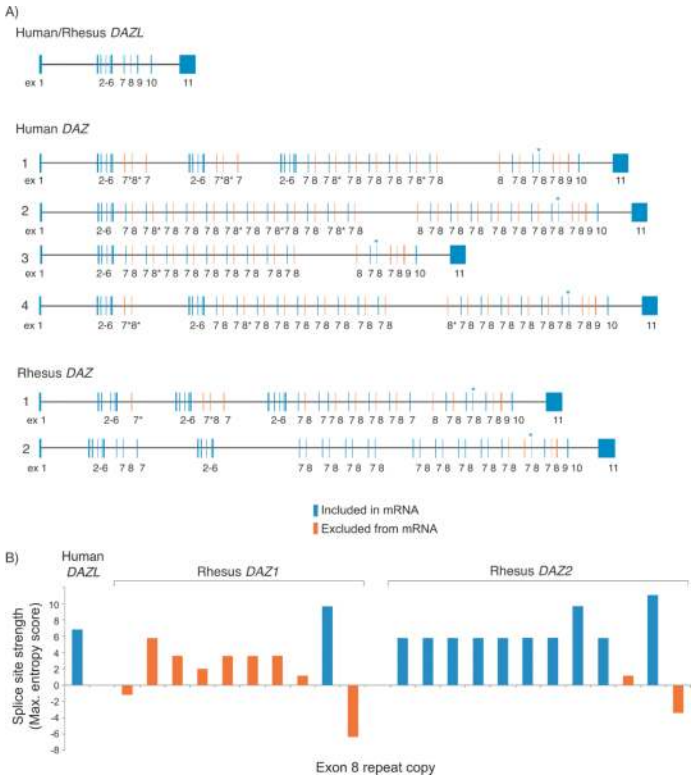

Figure 2.

Structure of members of the *DAZ* gene family in human and rhesus macaque. A: The exon-intron structure of each member of the *DAZ* gene family is shown to scale. Vertical lines and boxes represent exons, and exon number is indicated below. Horizontal lines represent introns. Exons that are spliced into mRNA product are shown in blue, and exons that are in genomic sequence but not spliced into mRNA are shown in orange. Asterisk after exon number indicates presence of mutation disrupting dinucleotide splice site; absence of asterisk indicates intact dinucleotide splice site. Asterisk above exon 8 indicates shortened exon. Splicing patterns for human [7,8] and rhesus [5] genes determined by cDNA sequencing. B: Donor splice site strength of various versions of exon 8 was evaluated using maximum entropy analysis [39] of 9-mers at 5' splice sites. Scores for amplified copies of exon 8 in rhesus *DAZ1* and *DAZ2* are shown. Score for human *DAZL* is shown at left. Bars in graph are color-coded to indicate inclusion or exclusion in mRNA splicing as shown.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

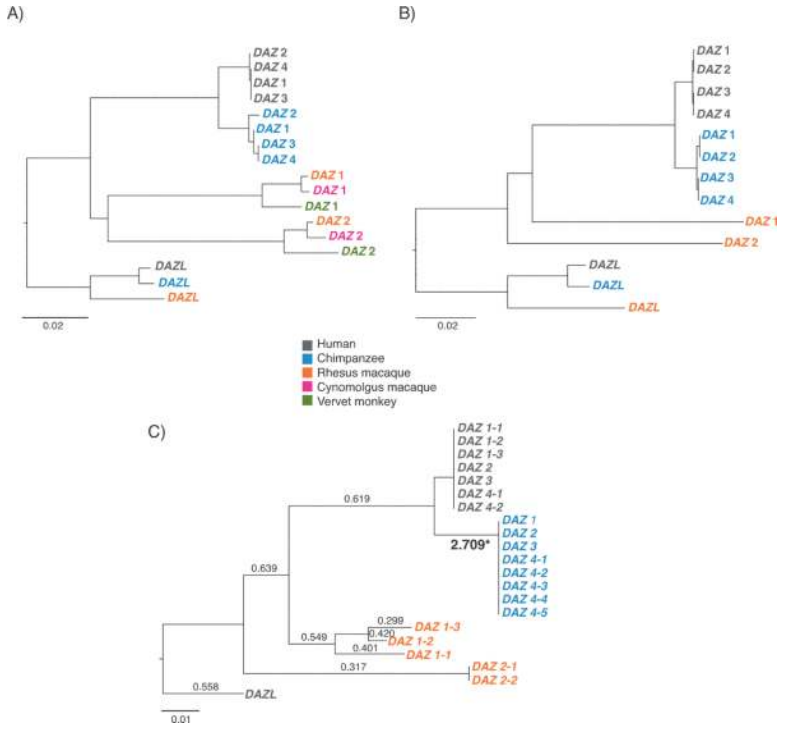

Figure 3.

Phylogenetic analyses of *DAZ* gene family members in primates and evidence for positive selection in chimpanzee. A–C: Maximum likelihood trees based on aligned sequences from (A) mRNAs, (B) introns, and (C) exons 2–6. Sequences were aligned using ClustalW and adjusted by hand in MacVector 12.0. Phylogenetic trees were generated using DNAML in Phylip 3.69 [48] with default parameters (http://cmgm.stanford.edu/phylip/dnaml.html). Graphical representations of trees were generated using FigTree 1.3.1 (http://tree.bio.ed.ac.uk/software/figtree). Numerical identifiers in *DAZ* gene names differentiate copy number only and do not indicate interspecies orthologies. Scale for branch lengths given in substitutions per site. C: For genes with multiple RRM repeats, repeat number is indicated after dash. Lineage-specific dN/dS ratios were calculated using the free-ratio model implemented in PAML [46]. All non-zero dN/dS ratios are indicated on branches. To determine statistical significance, the log-likelihood ratio test was used to compare observed ratio to a neutral evolution model where dN/dS is fixed at 1 (model 1 vs. 2 in PAML) [47].

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

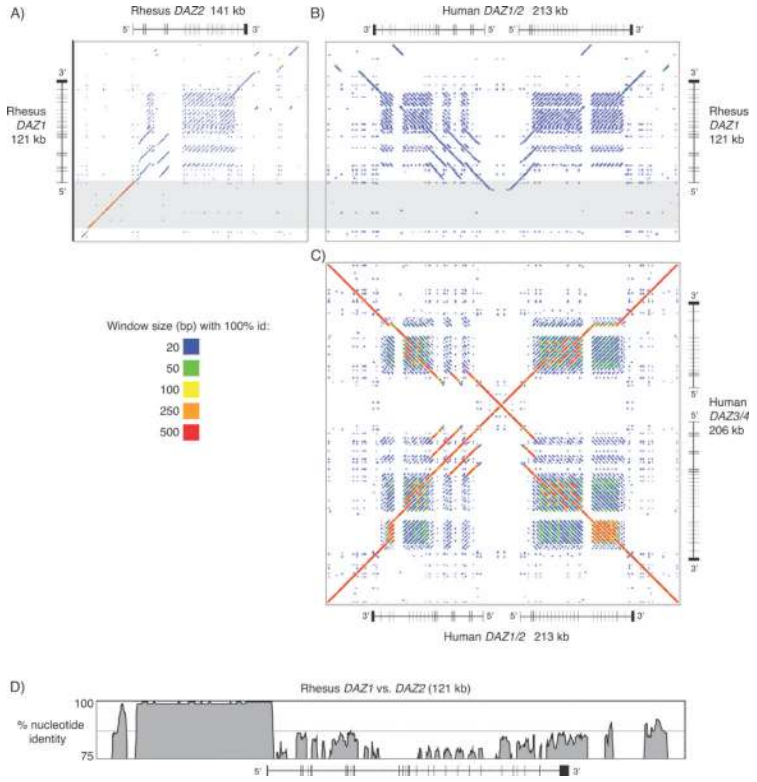

Figure 4.

Degree of conservation of *DAZ* genomic sequences within species and between species. A–C: Dot-plot analyses of (A) rhesus *DAZ1* vs. rhesus *DAZ2*, (B) rhesus *DAZ1* vs. human *DAZ1* and *DAZ2* (in human MSY palindrome P2) and (C) human *DAZ1* and *DAZ2* (palindrome P2) vs. human *DAZ3* and *DAZ4* (palindrome P1). Each dot within the plot is color-coded to reflect window size as indicated. Position, orientation, and exon-intron structure of genes are shown schematically on each axis. Gray shading indicates location of conserved 30-kb upstream region in rhesus, which is largely absent in human. D: Sliding-window analyses for *DAZ* genomic sequences in rhesus. Pairwise percent identity was calculated in 1000-bp windows and displayed using VISTA [49].

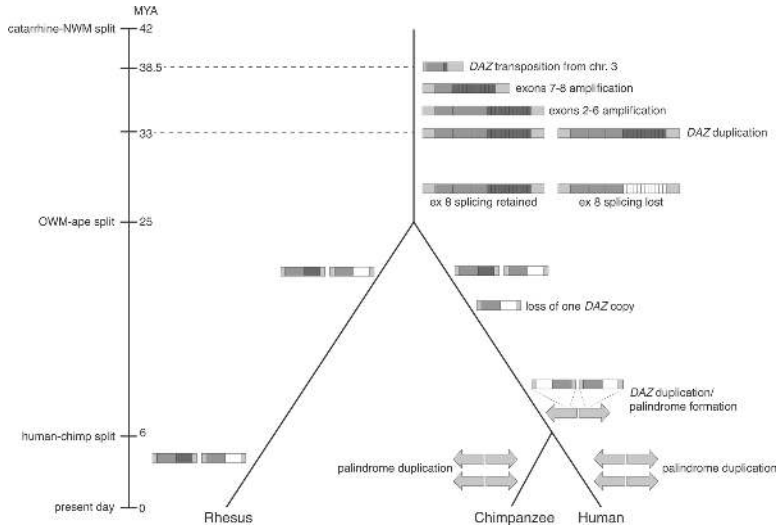

Figure 5.

Evolutionary history of the *DAZ* gene family. Phylogenetic tree shows evolutionary relationships between human, chimpanzee, and rhesus. Timeline is shown at left. Species divergence dates from http://www.timetree.org [25]. Dates of chromosome 3 transposition and initial *DAZ* duplication estimated as described in text. Timing of all other depicted events is only relative to other events. *DAZ* genes are shown as rectangles, with shading indicating major repeat domains. Palindromes shown as oppositely-facing block arrows. Abbreviations: MYA, millions of years ago; OWM, Old World monkey; NWM, New World monkey.

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript