SUPPLEMENTAL INFORMATION

VWF sequences

Human (*Homo sapiens*) VWF cDNA sequence was compiled from GenBank entries M10321.1 (1), M17588.1 (2) and X04385.1 (3). This full length cDNA sequence is the same as that of refseq NM 000552.3 except for a few known single nucleotide polymorphisms.

Human VWF gene sequence was based on AH005287.1, a segmented set of 38 entries M25828.1 through M25865.1 that covers all 52 exons (4), and M60675.1, which contains the sequence of a 21.3 kb gene fragment spanning exons 23-34 (5). The human VWF pseudogene sequence is M60676.1, which comprises a 21 kb pseudogene on chromosome 22 that corresponds to exons 23-34 of the VWF gene on chromosome 12 (5).

White-tufted-ear marmoset (*Callithrix jachus*) VWF cDNA sequence is predicted refseq XM_002752243.1. Chimpanzee (*Pan troglodytes*) VWF cDNA sequence is predicted refseq XM_508945.2. Mouse (*Mus musculus*) VWF cDNA sequence is refseq NM_011708.3. Giant panda (*Ailuropoda melanoleuca*) VWF cDNA sequence is predicted refseq XM_002930091.1.

VWF cDNA sequences for 25 species were assembled *de novo* from genomic sequences. Nucleotide sequence data reported for 23 species are available in the Third Party Annotation Section of the DDBJ/EMBL/GenBank databases under the accession numbers TPA: BK007980-BK008005. VWF cDNA sequences for western lowland gorilla (*Gorilla gorilla*) and bottlenosed dolphin (*Tursiops truncatus*) are appended below.

New VWF gene and cDNA sequences were identified by using human VWF genomic DNA, cDNA and protein sequences for iterative BLASTN and TBLASTN (NCBI version 2.2.24) searches of GenBank sequences including genome assemblies, high throughput genome sequences (htgs), genome survey sequences (gss), whole genome shotgun sequences (wgss), non-redundant nucleotide sequences, expressed sequence tags (ESTs), trace archives and short read archives. Exons and splice sites were identified based on sequence similarity using the programs SeqMan Pro, MegAlign, and EditSeq (version 8.1.5, DNASTAR, Inc., Madison, WI) and assembled into cDNA sequences. In some cases splice site prediction with the program SplicePort (version 1) (6) was used to support the identification of exon 2, which encodes a cleaved signal peptide. Remaining gaps were assigned lengths based on the conserved length of the corresponding exons for VWF from other species.

Rhesus monkey (*Macaca mulatta*) VWF cDNA sequence BK007989 is similar to predicted refseq XM_002807974.1, except that XM_002807974.1 lacks complete sequence for exon 28. BK007989 was derived from wgss AANU01190199.1 (exons 1-3), AANU01190200.1 (exons 4-5), AANU01190202.1 (exon 6), AANU01190205.1 (exons 7-18), AANU01190206.1 (exons 19-20), AANU01190207.1 (exons 21-28), AANU01190208.1 (exons 28-42), AANU01190209.1 (exons 43-47), and AANU01190210.1 (exons 48-52). Gaps and errors in exon 28 were corrected from traces TI490421287, TI852218215, and TI358561899.

Dog (*Canis lupus familiaris*) VWF cDNA sequence BK007981 is similar to predicted NM_001002932.1, except that parts of NM_001002932.1 are inconsistent with other cDNA entries or genomic traces for exons 7, 11, and 42. BK007981 was derived from wgss AAEX02011162.1.

Cattle (*Bos taurus*) VWF cDNA sequence BK007980 is derived from wgss AAFC03120372.1 (exons 1-3), AAFC03120374.1 (exons 6-7), and AAFC03028507.1 (exons 11-52). Gaps in this assembly were filled from traces TI419342196, TI515279818, TI425321040, TI589958794, and TI594837491. Exons

and splice sites were identified by comparison to cDNA sequences AW354654.1, BF777006.1, EE333226.1, BI535247.1, AW352873.1, BE749548.1, BM288253.1, AW658424.1, CK955532.1, and CK774294.1.

Pig (*Sus scrofa*) VWF cDNA sequence BK007995 is derived from htgs AC163958.2 and FP102760.2, which do not cover exons 1-5. These exons were obtained from cDNA sequences BP459041.1 and DT327670.1. Exons and splice sites were identified by comparison to cDNA sequences BE030430.1, BE236207.1, BM485003.1, BF442911.1, BF443270.1, BI346678.1, and BF703224.1.

Horse (*Equus caballus*) VWF cDNA sequence BK007985 is derived from wgss AAWR02029170.1 (exons 1-15), AAWR02029169.1 (exon 16), and AAWR02029168.1 (exons 18-52). Exon 17 was provided by trace TI1472575032.

Rat (*Rattus norvegicus*) VWF cDNA sequence BK007994 is similar to predicted refseq XM_342759.1, except that XM_342759.1 lacks exons 1 and 5. BK007994 is derived from wgss AAHX01032881.1 (exons 1-37), AAHX01032882.1 (exons 38-47), AAHX01032883.1 (exons 48-52). The majority of the sequence was confirmed and splice sites were identified by alignment with cDNAs FM070253.1, CB547010.1, CB736674.1, U50044.1, AJ224673.1, CB545392.1, CB610389.1, CB581251.1, CB703030.1, CF113897.1, and BI298314.1.

Rabbit (*Oryctolagus cuniculus*) VWF cDNA sequence BK007991 is derived from wgss AAGW02051559.1 (exons 1-5), AAGW02051558.1 (exons 6-20), AAGW02051557.1 (exons 21-38), AAGW02051556.1 (exons 39-43), and AAGW02051555.1 (exons 44-52). Errors at the 5' end of exon 34 were corrected from trace TI855913069. A frameshift in exon 28 was corrected from TI651097624. The splice site between exons 43-44 was identified by comparison to cDNA EC620075.1.

Chicken (*Gallus gallus*) VWF cDNA sequence BK007988 is derived from wgss AADN02065137.1 (exons 1-15), AADN02065136.1 (exons 16-21), AADN02065135.1 (exons 22-30), AADN02065133.1 (exons 31-34), and AADN02065132.1 (exons 35-52). Identification of exons 1-4, 14-17, 25-28, 29-31, and 38-52 is supported by comparison to cDNAs BU323363.1, BU462510.1, BU331027.1, BU309203.1, BQ037193.2, BM485548.1, AI979961.1, BM426048.1, BX258791.1, BM488796.1, and BX258790.1.

Zebra finch (*Taeniopygia guttata*) VWF cDNA sequence BK007997 is derived from wgss ABQF01036128.1 (exons 1-27), ABQF01036130.1 (exons 28-30), and ABQF01036131.1 (exons 31-52). The 5' end of exon 28 was completed from trace TI1261325877. The sequence is similar to predicted refseq XM_002188179.1, except that exons 2, 23, 24 and 28 in XM_002188179.1 required correction based on alignments with corresponding chicken, anole, human and frog sequences.

Western clawed frog (*Xenopus tropicalis*) VWF cDNA sequence BK007996 is derived from wgss AAMC01011472.1 (exon 2), AAMC01011474.1 (exons 3-6), AAMC01011475.1 (exons 7-32), AAMC01011476.1 (exons 33-40), AAMC01011477.1 (exons 41-42), AAMC01011478.1 (exons 43-44), and AAMC01011479.1 (exons 45-52). Splice sites for exons 28-34 and exons 40-52 were identified by comparison to ESTs DR835756.1, DN007177.1, DN080731.1, EG653390.1, EL715901.1, DN083600.1, and EL715900.1. BK007996 is similar to predicted refseq XM_002932296.1 except that XM_002932296.1 has errors in exons 2 and 33.

Sumatran orangutan (*Pongo abelii*) VWF cDNA sequence BK007993 is derived from wgss ABGA01401196.1 (exons 1-3), ABGA01275753.1 (exons 1-3), ABGA01268156.1 (exons 4-5), ABGA01268154.1 (exon 6), ABGA01063082.1 (exons 7-15), ABGA01063080.1 (exons 16-18), ABGA01063078.1 (exons 19-20), ABGA01063077.1 (exons 21-34), ABGA01171174.1 (exons 35-37), ABGA01324188.1 (exons 38), ABGA01171175.1 (exons 39-40), ABGA01171176.1 (exons 41-42),

ABGA01013475.1 (exons 43), ABGA01013476.1 (exons 44-47), and ABGA01090113.1 (exons 48-52). A frameshift in exon 14 was corrected from trace TI897102100.

African elephant (*Loxodonta africana*) VWF cDNA sequences BK008002 (exons 1-17) and BK008003 (exons 18-52) are derived from wgss AAGU03036114.1 (exons 1-16), AAGU03036113.1 (exons 18-47), and AAGU03036112.1 (exons 48-52). Exon 17 was provided by trace TI1744370414. The gap between BK008002 and BK008003 corresponds to approximately 46 nucleotides at the 5' end of exon 18.

Domestic cat (*Felis catus*) VWF cDNA sequence BK007986 is derived from wgss ACBE01225964.1 (exon 1), ACBE01225963.1 (exons 3-4), ACBE01225962.1 (exon 5), ACBE01225960.1 (exon 6), ACBE01225961.1 (exon 7), ACBE01225955.1 (exons 8-15), ACBE01225954.1 (exon 16), ACBE01225952.1 (exon 17), ACBE01225951.1 (exon 18), ACBE01225949.1 (exons 19-22), ACBE01225948.1 (exons 23-27), ACBE01225947.1 (exons 28-34), ACBE01225946.1 (exon 35), ACBE01225945.1 (exons 36-37), ACBE01225944.1 (exon 38), ACBE01225943.1 (exons 39-42), ACBE01225942.1 (exons 43-45), ACBE01225941.1 (exons 46-47), ACBE01225940.1 (exon 48), and ACBE01225939.1 (exons 49-52). Exon 2 was provided by trace TI1950120789. The 3' end of exon 4 was provided by trace TI898344027.

Large flying fox (*Pteropus vampyrus*) VWF cDNA sequence BK008004 (exons 1-20) and BK008005 (exons 21-52) are derived from wgss ABRP01038654.1 (exon 2), ABRP01360519.1 (exon 3), ABRP01038656.1 (exons 4-6), ABRP01038659.1 (exons 7-10), ABRP01038660.1 (exons 11-12), ABRP01038661.1 (exon 13), ABRP01038662.1 (exons 14-18), ABRP01038663.1 (exons 19-20), ABRP01038664.1 (exon 21), ABRP01038665.1 (exons 22-24), ABRP01038666.1 (exons 25-34), ABRP01038668.1 (exons 35-40), ABRP01038669.1 (exons 41-47), and ABRP01038672.1 (exons 48-52). Exon 12 was edited based on trace TI1379389256. The gap between BK008004 and BK008005 corresponds to approximately 56 nt at the 3' end of exon 20.

Nine-banded armadillo (*Dasypus novemcinctus*) VWF cDNA sequence BK007983 is derived from htgs AC163260.2 (exons 1-41); wgss AAGV020435885.1 (exon 42), AAGV020435888.1 (exon 44), AAGV020435889.1 (exons 46-47), AAGV020561273.1 (exon 51); and traces TI1895851916 (exon 17), TI1885723830 (exon 43), TI1873705844 (exon 45), TI2005246912 (exon 48), TI1641417980 (exon 49), TI1803631052 (exon 50), and TI1873812382 (exon 52).

Domestic guinea pig (*Cavia porcellus*) VWF cDNA sequence BK007982 is derived from wgss AAKN02051671.1 (exons 1-16) and AAKN02051670.1 (exons 17-52).

Gray short-tailed opossum (*Monodelphis domestica*) VWF cDNA sequence BK007990 is derived from wgss AAFR03037682.1 (exons 2-6), AAFR03037678.1 (exons 7-37), AAFR03037676.1 (exons 38- 39), and AAFR03037675.1 (exons 40-52).

Green anole lizard (*Anolis carolensis*) VWF cDNA sequences BK008000 (exons 1-32) and BK008001 (exons 34-52) are derived from wgss AAWZ02019413.1 (exons 1-14), AAWZ02019412.1 (exons 15-19), AAWZ02019410.1 (exons 22-31), AAWZ02019409.1 (exon 32), AAWZ02019408.1 (exons 34-46), and AAWZ02019407.1 (exons 47-52). The gap between BK008000 and BK008001 corresponds to exon 33, or approximately 44 nucleotides.

Zebrafish (*Danio rerio*) VWF cDNA sequence BK007984 is derived from wgss CABZ01015114.1 (exon 2), CABZ01015113.1 (exons 3-7), CABZ01015112.1 (exons 8-18), CABZ01015111.1 (exons 19- 20), CABZ01015110.1 (exon 21), CABZ01015109.1 (exons 22-33), CABZ01015108.1 (exons 34-36), CABZ01015107.1 (exons 37-45), and CABZ01015106.1 (exons 46-52). Splice sites for exons 38-52 are supported by comparison to cDNA entries EB902243.1, EH543632.1, and EB978371.1 BK007984 is similar to predicted XM 002665310.1 , except that exons 4, 5, 7, 8, 10, 29, 43, and 44 appear to be incorrect in XM_002665310.1.

Japanese medaka (*Oryzias latipes*) VWF cDNA sequence BK007992 is derived from wgss BAAF04048397.1 (exons 2-51) and BAAE01258935.1 (exons 45-52).

Three-spined stickleback (*Gasterosteus aculeatus*) VWF cDNA sequence BK007987 is derived from wgss AANH01001674.1 (exons 2-52).

Spotted green pufferfish (*Tetraodon nigroviridus*) VWF cDNA sequence BK007998 is derived from wgss CAEE01015122.1 (exons 2-52).

Torafugu (*Takifugu rubripes*) VWF cDNA sequence BK007999 is derived from wgss CAAB01000205.1 (exons 2-52).

Western lowland gorilla (*Gorilla gorilla*) VWF cDNA sequence (VWFgorilla, appended after Table S1) is derived from wgss CABD02074492.1 (exons 1-3), CABD02074490.1 (exons 4-5), CABD02074487.1 (exon 6), CABD02074482.1 (exons 7-10), CABD02074481.1 (exons 11-16), CABD02074479.1 (exon 17), CABD02074478.1 (exon 18), CABD02074477.1 (exons 19-24), CABD02230032.1 (exons 25-32), CABD02074475.1 (exons 33-34), CABD02074472.1 (exons 35-42), CABD02074469.1 (exons 43-47), and CABD02074467.1 (exons 48-52). Human, chimpanzee, gorilla and orangutan genomes contain a VWF locus on chromosome 12 and a partial unprocessed pseudogene on chromosome 22 that is approximately 96% identical to the gene sequence (5). The sequence of gorilla VWF exon 28 was confirmed to be from the VWF gene by sequencing of genomic DNA isolated from peripheral blood of a gorilla, using PCR primers to selectively amplify and sequence exon 28 from the VWF gene (5). Comparisons with the human VWF gene and pseudogene suggests that the sequences for gorilla VWF exons 25, 26, 29, 30, 31 and 33 are from the VWF gene, but that exons 23, 24, 27, 32, and 34 contain some pseudogene sequences, probably due to errors in genome assembly.

Bottlenosed dolphin (*Tursiops truncatus*) VWF cDNA sequence (VWFtursiops, appended after Table S1) is derived from wgss ABRN01165339.1 (exon 3), ABRN01165340.1 (exon 4), ABRN01165341.1 (exon 5), ABRN01165342.1 (exons 6-10), ABRN01165343.1 (exons 11-16), ABRN01165344.1 (exons 17-18), ABRN01165346.1 (exons 19-20), ABRN01165347.1 (exon 21), ABRN01165348.1 (exons 22-26), ABRN01165349.1 (exons 27-28), ABRN01165350.1 (exons 28-30), ABRN01165351.1 (exons 32-43), ABRN01165352.1 (exons 44-47), ABRN01165353.1 (exons 49-50), and ABRN01355946.1 (exons 51- 52). For alignments, gaps corresponding to exons 2, 31, and 48, as well as incomplete portions of exons 17 and 28, were assigned lengths equal to the corresponding human VWF exons.

Mucin sequences

Chicken (*Gallus gallus*) MUC2 cDNA sequence is based on predicted refseq XM_421035.2, which was edited to correct splice sites between exons 4 and 5 by alignment with chromosome reference assembly NC_006092.2 and other MUC2 homologs. MUC6 is predicted refseq XM_426405.2.

Dog (*Canis lupus familiaris*) MUC2 cDNA sequence is based on predicted refseq XM_540774.2. MUC5AC cDNA sequence is predicted refseq XM 540775.2, which is incorrectly annotated as MUC5B. MUC5B is predicted refseq XM_540777.2. MUC6 is predicted refseq XM_540773.2. MUC19 is predicted refseq XM_847204.1.

Cattle (*Bos taurus*) MUC19 cDNA sequence is predicted refseq XM_002687376.1.

Pig (*Sus scrofa*) MUC19 cDNA sequence is AF005273.1.

Horse (*Equus caballus*) MUC2 cDNA sequence is based on predicted refseq XM_001492997.1 (nt 1- 3673) and predicted refseq XM_001502065.1 (nt 3392-4233). MUC5AC is based on predicted refseq XR 044663.1, edited to correct exon 12 from reference assembly NC 009155.2 and genomic traces.

Human (*Homo sapiens*) MUC2 cDNA sequence is refseq NM_002457.2. MUC5AC is predicted refseq XM_001717880.2. MUC5B is predicted refseq XM_001129427.1. MUC6 is validated refseq NM 005961.2. MUC19 is predicted refseq XM 497341.3. This version has some errors in gene prediction at the 5' end but the region corresponding to domains D1D2D'D3 appears to be correct. The later version XM 497341.4 is missing substantial segments of domain D1.

Mouse (*Mus musculus*) MUC2 cDNA sequence is based on predicted refseq XM_001001905.1 (nt 1- 3375), EST BF578645.1 (nt 3376-3506), and predicted refseq XM_001001905.1 (nt 3507-7374). MUC5AC is based on validated refseq NM_010844.1 edited to add 3 nt at the 3' end of exon 10 based on alignment with other MUC5AC homologs and genomic working draft sequence AC020817.3. MUC5B is validated refseq NM_028801.2. MUC6 is cDNA BC120907.2. MUC19 is provisional refseq NM_207243.2.

Rat (*Rattus norvegicus*) MUC2 cDNA sequence is cDNA U07615.1. MUC5AC is partial cDNA AB042530.1. MUC5B is predicted refseq XM_238988.5. MUC6 is based on predicted refseq XM_215127.6, with editing of exons 2 and 5 based on alignment with mouse MUC6 cDNA BC120907.2 and genomic scaffold NW_047563.2. MUC19 is predicted sequence BK005555.1.

African clawed frog (*Xenopus laevis*) integumentary mucin B1 (FIMB1) cDNA sequence is Y08296.1.

Phylogenetic Analysis

Protein sequences of VWF and mucins were aligned with MegAlign (version 8.1.5, DNASTAR, Inc., Madison, WI) using the Clustal W method, Gonnet protein weight matrices, gap penalty 10, and gap length penalty 0.2. Because the proteins compared are relatively similar, the alignment is relatively insensitive to the choice of parameters or weight matrix. For example, nearly identical results were obtained with PAM or BLOSUM matrices, particularly with respect to alignment of Cys or His residues.

Species	Intron	Donor Sequence	Acceptor Sequence	Evidence
M. domestica	13	ggtgaagGCaagtcc	ctcctcAGttgtctc	Alignment with other VWF genes
A. carolinensis	26	cctgcagGCaagcat	tgaaacAGgaaaaat	Alignment with other VWF genes and A. carolinensis spliced mRNA transcripts
T. nigroviridus	39	attaaagGCaagctt	gtgtgcAGctcctgt	Alignment with other VWF genes
T. nigroviridus	48	tgtacagGCaagcaa	tcctgcAGgtgaacg	Alignment with other VWF genes
<i>T. rubripes</i>	39	attaaagGCaagctt	qtatacAGctcctgt	Alignment with other VWF genes
G. aculeatus	39	gtcagagGCaagatt	gtgtgcAGctccagt	Alignment with other VWF genes
O. latipes	4	aaaaaagGCacaagc	tctcgcAGactcact	Alignment with other VWF genes
D. rerio	44	gaggcagGCgagtgt	gtttgaAGgtcggag	Alignment with other VWF genes and <i>D. rerio</i> spliced mRNA transcripts

TABLE S1. **Non-canonical splice sites in VWF genes**

>VWFgorilla Gorilla gorilla VWF cDNA sequence agctcacagctattgtggtgggaaagggagggtggttggtggatgtcacagcttggcctttatctcccccagcagtgggg actccacagcccctgggctacataacagcaagacagtccggagctgtagcagacctgattgagcctttgcagcagctgag agcatggcctagggtgggcggcaccattgtccagcagctgagtttcccagggaccttggagatacccgcagccctcattt gcaggggaagatgattcctgccagagttgccggggtgctgcttgctctggccctcattttgccagggaccctttgtgcag aaggaactcgcggcaggtcatccatggcccgatgcagcctcttcggaaatgacttcgtcaacacctttgatgggagcatg tacagctttgcgggatactgcagttacctcctggcagggggctgccagaaacgctccttctcgattattggggacttcca gaatggcaagagagtgagtctctccgtgtatcttggggaattttttgacatccatttgtttgtcaatggtactgtgacac agggggaccaaagagtctccatgccctatgcctccaaagggctgtatctagaaactgaggctgggtactacaagctgtct ggtgaggcctatggctttgtggccaggatcgatggcagcggcaactttcaagtcctgctgtcagacagatacttcaacaa gacctgcgggctgtgtggcaactttaacatctttgctgaagatgactttatgacccaagaagggaccttgacctcggacc cttatgactttgccaactcatgggctctcagcagtggagaacagtggtgtgaacgggtatctcctcccagcagctcatgc aacatctcctctggggaaatgcagaagggcctgtgggagcagtgccagcttctgaagagcacctcggtgtttgcccgctg ccaccctctggtggaccccgagccttttgtggccctgtgtgagaagacgttgtgtgagtgtgctggggggctggagtgcg cctgccctgccttcctggagtacgcccggacctgtgcccaggagggaatggtgctgtacggctggaccgaccacagcgcg tgcagcccagtgtgccctgctggtatggagtataagcagtgtgtgtccccttgcgccaggacctgccagagcctgcacgt caatgaaatgtgtcaggagcgatgcgtggatggctgcagctgccctgagggacagctcctggatgaaggcctctgcgtgg agagcaccgagtgtccctgcgtgcattccggaaaacgctaccctcccggcgcctccctctctcgagactgcaacacctgc atttgccgaaacagccagtggatctgcagcaatgaagaatgtccaggggagtgcctcgtcacaggtcaatcccacttcaa gagctttgacaacagatacttcaccttcagtgggatttgccagtacctgctggcccgggattgccaggaccactccttct ccattgtcattgagaccgtccagtgtgctgatgaccgcgacgctgtgtgcacccgctccgtcaccctccggctgcctggc ctgcacaacagcctagtgaaactgaagcatgggggaggagttgccatggatggccaggacgtccagctccccctcctgaa aggtgacctccgcatccagcatacagtgacggcctccgtgcgcctcagctacggggaggacctgcagatggactgggatg gccgcgggaggctgctggtgaagctgtcccccgtctacgccgggaagacctgcggcctgtgtgggaattacaatggcaac cagggcgacgacttccttaccccctctgggctggcggagccccgggtggaggacttcgggaacgcctggaagctgcacgg ggactgccaggacctgcagaagcagcacagcgatccctgcgccctcaacccgcgcatgaccaggttctccgaggaggcgt gcgcggtcctgacgtcccccacattcgaggcctgccatcgtgccgtcagcccgctgccctacctgcggaactgccgctac gacgtgtgctcctgctcggacggccgcgagtgcctgtgcggcgccctggccagctatgccgcggcctgcgcggggagagg cgtgcgcgtcgcgtggcgcgagccaggccgctgtgagctgaactgcccgaaaggccaggtgtacctgcagtgcgggaccc cctgcaacctgacctgccgctctctctcttacccggatgaggaatgcaatgaggcctgcctggagggctgcttctgcccc ccagggctctacatggatgagaggggggactgcgtgcccaaggcccagtgcccctgttactatgatggtgagatcttcca gccagaagacatcttctcagaccatcacaccatgtgctactgtgaggatggcttcatgcactgtaccatgagtggagtcc ccggaagcttgctgcccgacgctgtcctcagcagtcccctgtctcatcgcagcaaaaggagcctatcctgtcggcccccc atggtcaagctggtgtgtcccgctgacaacctgcgggctgaagggctcgagtgtgccaaaacgtgccagaattacgacct ggagtgcatgagcatgggctgtgtctctggctgcctctgccccccgggcatggtccggcatgagaacagatgtgtggccc tggaaaggtgtccctgcttccatcagggcaaggagtatgcccctggagaaacagtgaagattggctgcaacacttgtgtc tgtcgggaccggaagtggaactgcacagaccatgtgtgtgacgccacgtgctccacgatcggcatggcccactacctcac cttcgacgggctcaaatacatgttccccggggagtgccagtacgttctggtgcaggattactgcggcagtaaccctggga cctttcggatcctagtggggaacgagggatgcagccatccctcagtgaaatgcaagaaacgggtcaccatcctggtggag ggaggagagattgagctgttcgacggggaggcgaatgtgaagaggcccatgaaggatgagactcactttgaggtggtgga gtctggccggtacatcattctgctgctgggcaaagccctctccgtggtctgggaccgccacctgagcatctccgtggtcc tgaagcagacataccaggagaaagtgtgtggcctgtgtgggaattttgatggcatccagaacaatgacctcaccagcagc aacctccaagtggaggaagaccctgtggactttgggaactcctggagagtgagctcgcagtgtgctgacaccagaaaagt gcctctggactcatcccctgccacctgccataacaacatcatgaagcagacaatggtggattcctcctgtagaatcctta ccagtgacgtcttccaggactgcaacaagctggtggaccccgagccatacctggacgtctgcatttacgacacatgctcc tgtgagtccattggggactgcgcctgcttctgcgacaccattgctgcctatgcccacgtgtgtgcccagcatggcaaggt ggtgacctggaggacggccacattgtgcccccagagctgcgaggagaggaatctccgggagaacgggtatgagtgcgagt ggcgctataacagctgtgcacctgcctgtcaagtcacgtgccagcaccctgagccgctggcctgccctgtgcagtgtgtg gagggctgccatgcccactgccctccagggaaaatcctggatgagcttttgcagacctgcgttgaccctgaagactgccc ggtgtgtgaggtggctggctggtgttttgcctcaaaaaagaaagtcaccttgaatcccagtgaccctgagcactgccaga tttgccactgtgatggtgtcaacctcacctgtgaagcctgccaggagccgggaggcctggtggggcctcccacagatgcc ccggtgagccccaccactccgtatgtggaggacatcttggaaccgccgttgcacgatttctactgcagcaggctactgga cctggtcttcctgctggatggctcctccaggctgtccgaggctgagtttgaagtgctgaaggcctttgtggtggacatga tggagcggctgcgcatctcccagaagtgggtacgcgtggccgtggtggagtaccacgacggctcccacgcctacattggg ctcaaggaccggaagcgaccgtcagagctgcggcgcattgccagccaggtgaagtatgcgggcagccaggtggcctccac cagcgaggtcttgaaatacacactgttccaaatcttcagcaagatcgaccgccctgaagcctcccgcatcgccctgctcc tgatggccagccaggagccccagcggatgtcccggaactttgtccgctacgtccagggcctgaagaagaagaaggtcatc gtgatcccagtgggcattgggccccatgccaacctcaagcagatccgcctcattgagaagcaggcccctgagaacaaggc cttcgtgctgagtggtgtggatgagctggagcagcaaagggacgagatcgttagctacctctgtgaccttgcccctgaag cccctcctcctactctgcccccccacatggcacaagtcactgtgggcccggggctcttgggggtttcgaccctggggccc aagaggaactccatggttctggatgtggcgtttgtcctggaaggatcggacaaaattggtgaagccgacttcaacaggag caaggagttcatggaggaggtgattcagcggatggatgtgggccaggacagcatccatgtcacggtgctgcagtactcct acatggtgaccgtggagtaccccttcagcgaggcacagtccaaaggggacatcctgcagcgggtgcgagagatccgctac cagggcggcaacaggaccaacactgggctggccctgaggtacctctctgaccacagcttcttggtcagccagggtgaccg ggagcaggcgcccaacctggtctacatggtcaccggaaatcctgcctctgatgagatcaagaggctgcctggagacatcc aggtggtgcccattggagtgggccctaatgccaacatgcaggagctggagaggatcggctggcccaatgcccccatcctc atccaggactttgagacgctcccccgagaggctcctgacctggtgctgcagaggtgctgctccggagaggggctgcagat ccccaccctctcccctgcatctgactgcagccagcccctggacgtgatccttctcctggatggctcctccagtttcccag cttcttattttgatgaaatgaagagtttcgccaaggctttcatttcagaagccaatatagggcctcgtctcactcaggtg tcagtgctgcagtatggaagcatcaccaccattgacgtgccatggaacgtggtcccggagaaagcccatttgctgagcct tgtggacgtcatgcagcgggagggaggccccagccaaatcggggatgccttgggctttgctttgcgatacttgacttcag aaatgcatggtgccaggccgggagcctcgaaggcggtggtcatcctggtcatggatgtctctgtggattcagtggatgca tcagctgatgccgccaggtccaacagagtgacagtgttccctattggaattggagatcactacgatgcagcccagctacg gatgttggccggcccagcaggcgactccaacatggtgaagctccagcaaattgaagacctccctaccatggtcaccttgg gcaattccttcctccacaaactgtgctctggatttgttaggatttgcatggatgaggatgggaatgagaagaggcctggg gacgtctggaccttgccagaccagtgccacaccgtgacttgccagccagatggccagaccttgctgaagagtcatcgggt caactgtgaccggggaccgaggccttcatgccctaacagccagtcccctgttaaagtggaagagacctgtggctgccgct ggacctgcccctgcgtgtgcacaggcagctccactcggcacatcgtgacctttgatgggcagaatttcaagctgactggc agctgttcttatgtcctatttcaaaacaaggagcaggacctggaggtgattctccataatggtccctgcagccctggagc aaggcagggctgcatgaaatccatcgaggtgaagcacagtgccctctccgtcgagctgcacagtgacatggaggtgacgg tgaatgggagactggtctctgttccttacgtgggtgggaacatggaagtcaacgtttatggtgccatcatgcatgaggtc agattcaatcaccttggtcacatcttcacattcactccacaaaacaatgagttccaactgcagctcagccccaagacttt tgcttcaaagatgtatggtctgtgtgggatctgtgatgagaatggagccaatgacttcatgctgagggatggcacagtca ccacagactggaaaacacttgttcaggaatggactgtgcaacggccagggcagacgtgccagcccatcctggaggagcag tgtcttgtccctgacagctcccactgccaggtcctcctcttaccactgtttgctgaatgccacaaggtcctggctccagc cacattctatgccatctgccagcaggacagttgccaccaggagcgagtgtgtgaggtgatcgcctcttatgcccacctct gtcggaccaatggggtctgcattgactggaggacgcctgatttctgtgctatgtcatgcccaccatctctggtctacaac cactgtgagcatggctgtccccggcactgtgatggcaacgtgagctcctgtggggaccatccctccgaaggctgtttctg ccctccaaataaagtcatgttggaaggcagctgtgtccctgaagaggcctgcactcagtgcattggtgaggatggagtcc agcaccagttcctggaagcctgggtcccggaccaccagccctgtcagatctgcacgtgcctcagcgggcggaaggtcaac tgcacaacgcagccctgccccacggccaaagctcccacgtgtggcctgtgtgaagtagcccacctccgccagaatgcaga ccagtgctgccccgagtatgagtgtgtgtgtgacccagtgagctgtgacctgcccccagtgcctcactgtgaacgtggcc tccagcccacactgaccaaccctggcgagtgcagacccaacttcacctgcgcctgcaggaaggaggagtgcaaaagagtg tccccaccctcctgccccccgcaccgtttgcccacccttcggaagacccagtgctgtgatgagtatgagtgtgcctgcaa ctgtgtcaactccacagtgagctgtccccttgggtacttggcctcaaccgccaccaatgactgtggctgtaccacaacca cctgccttcctgacaaggtgtgtgtccaccgaagcaccatctaccctgtgggccagttctgggaggagggctgcgatgtg tgcacctgcaccgacatggaggatgtcgtgatgggcctccgcgtggcccagtgctcccagaagccctgtgaggacagctg tcggtcgggcttcacttacgttctgcatgaaggcgagtgctgtggaaggtgcctgccatctgcctgtgaggtggtgaccg gctcaccgcggggggactcccagtcttcctggaagagtgtcggctcccagtgggcctccccggagaacccctgcctcatc agtgagtgtgtccgagtgaaggaggaggtctttatacaacaaaggaacgtctcctgcccccagttggaggtccctgtctg cccctcgggctttcagctgagctgtaagacctcagcgtgctgcccaagctgtcgctgtgagcgcatggaggcctgcatgc tcaatggcactgtcattgggcccgggaagactgtgatgatcgatgtgtgcacgacctgccgctgcatggtgcaggtgggg gtcatctctggattcaagctggagtgcaggaagaccacctgcaacccctgccccctgggttacaaggaagaaaataacac aggtgaatgttgtgggagatgtttgcctacggcttgcaccattcagctaagaggaggacagatcatgacactgaagcgtg atgagacgctccaggatggctgtgatactcacttctgcaaggtcaatgagagaggagagtacttctgggagaagagggtc acaggctgcccaccctttgatgaacacaagtgtctggctgagggaggtaaaattatgaaaattccaggcacctgctgtga

cacatgtgaggagcctgagtgcaacgacatcactgccaggctgcagtatgtcaaggtgggaagctgtaagtctgaagtag aggtggatatccactactgccagggcaaatgtgccagcaaagccatgtactccattgacatcaacgatgtgcaggaccag tgctcctgctgttctccgacacggacggagcccatgcaggtggccctgcactgcaccaatggctctgttgtgtaccatga ggttctcaatgccatggagtgcaaatgctcccccaggaagtgcagcaagtgaggctgctgcacctgcatgggtgcctgct gctgcctgccttggactgacggccaggccagagtgctgccagtcctctgcatgttctgctcttgtgcccttctgagccca caataaaggctgagctcttatcttgc

>VWFtursiops Tursiops truncatus VWF cDNA sequence

atgnngggccctttgtggagaaggcgctct tggcaggtcgtcgatggcccgatgcagcctcttcggaggtgattccatcaacacctttgatgagagcatgtacagctttg agggaagctgtagttatctcctggcgggggactgccagacacactccttctctctcataggggacttccaagatggcaag agagtgagcctttctgtgtatcttggggaatttttcgatatccatgtgtttgtcaatggtactgtgctgcagggggacca gagggtctccatgccctatacctccaaagggctgtacctagaaactgaggctgggtactacaagctgtccagcgaggcct acggctttgtggccaggattgatggcagcggcagctttcaagtcctgctgtcagacagacacttcaacaagacctgtggg ctgtgcggcaactttaacatctttgccgaagatgacttccggactcaagaagggaccttgacctcggacccctacgactt tgccaactcctgggccctgagcagtggggaacagcggtgccagcgggcgtcccctcccagcagcgtgtgcaatgtctcct ccgagctgcagaagggcctgtgggagcagtgccagcttctgaagagcgcctcggtgttcgcccgctgccaccccctggtg gaccccgagcctttcgtggccctgtgtgagaggatgctgtgcacgtgtgaccaggggctgctgtgcgcctgccctgtgct cctggagtacgcccgggcctgcgcccagcaggggatgctgctgtacggctggacggaccacagctcctgccgaccggact gccccgcgggcatggagtacaaggcatgcgtgtccccctgccccagagactgccggagcctgcacgtcagtgaagtgtgt cagcagcagtgcgtggatggctgcagctgccctgagggacagttccttgacgaaggccgctgtgtgaaaagtgccgagtg ttcctgcgtacattctgggaagcggtaccctccgggcgcctcgctctcgagagactgcaacacctgcatatgccgaaaca gccagtgggtctgcagcaacgaagagtgtccaggggcgtgtctcatcacaggacaatcccacttcaagagctttgacaac aggcacttcaccttcagtgggatctgccagtacctgctggcccgggactgccaggaccacaccttctccatcatcataga gacggtccagtgtgccgacgaccccgacgcggtctgcacgcgctcggtcaccgtccgtctgcccggccagcaccgcaacg tcgtgaagctgaagcacgggggaggcgttgccgtggacggtcaggacatccagatgcccctcctgcaaggtgacctccgc atccagcacaccgtgatggcctccgtgcacctcagctacggggaggacctgcagatagactgggacggccgcgggaggct gctggtgaagctgtccccggtctactcggggaggacgtgcggcctgtgcgggaattacaacggcaaccagggggacgact ttctgacgcccgcgggcctggtggagcccctggtggagcacttcggaaactcctggaagctgcgtgcggactgcgaggac ctgcagaagcagcccagcgacccctgcagcctcaacccgcgcctgaccagattcgcggacgagtcctgcgccatcctgac gtcgcccaagttcaaggcctgccacggcaccgtgggcccgctgccctacctgcagaactgccgctacgacgtgtgctcct gcaccaacggcagagactgcctgtgcgacgcggtggccaactacgcggcggcctgtgcccggagcggcgtgcgcatcggc tggcgggagcccagcttctgcgcgctgagctgcccacgaggccaggtgtacctgcagtgtgggaccccctgcaacctcac ctgccgctccctctcttacccggacgaggaatgcactgaggtctgcctggagggctgcttctgccccctggggctctacc tggatgagaccggggcctgtgtgcccaaggcccagtgcccctgttactacgacggcgagatcttccagcccgaagacgtc ttctcggaccatcacaccatgtgctactgtgaggatggcttcatgcactgctccacgagtggagccccggggagcctgct gcccgactcggtcctcagcagccccctgtcccaccgcagcaaaagaagcctgtcctgccagccccccatggtcaagctgg tgtgtcctgctgacagcccgaaggccgaagggctggagtgcgccaaaacctgccagaactacgacctggagtgcgtgagc atgggctgcgtgtccggctgcctctgccccccgggcatggtccggcaggacaacaggtgtgtggccctggaaaggtgtcc ctgcttccaccagggcagagagtacagacccggagagacggtgaaggtggactgcaatacttgtgtctgtcgggaccgga agtggaactgcacggaccacgtgtgcgatgccacctgctccgccctgggcctggctcactacctcacctttgacgggctc aagtacctgtttcccggggaatgccagtacgtcctggcacaggattactgtggcagtaaccctgggaccttccggatcct ggtggggaacgaggggtgcggctacccctccctgaaatgcaggaggcgcgtcaccatcctggtggaaggaggcgagattg agctgtttgatggggaggtgagcgtgaagaggtccatgaaggatgagacttactttgaggtggtggagtccggcctgtac atcactgtgctgctgggcaaggccctctccgtggtctgggaccgccgcctgggtatctctgtgttcctgaagcagactta ccaggagcaggtgtgtggcctgtgtgggaactttgacggcatccagaacaacgacctcaccagcagtagcctccaagtgg aggaagaccccgtggacttcgggaattcctggaaagtgagcccgcggtgcgccgacacccagaaagtgccactggacccc tcccctgccacctgccacaacaacgtcctgaagcagacgatggtggactcctcctgcagcatcctcaccagcgacatctt ccaggagtgcaacaggctggtggaccccgagccatacctggatgtttgcatctacgacacctgttcctgcgagtccgtcg gggactgtgcctgcttctgcgacgccatcgctgcctacgcccacgtgtgtgcccggcacggcgaggtggtggcctggagg acgcccacgttgtgcccccagagctgcgaggaacggaacctgcgggacagtgggtaccagtgtgagtggcgttacaacag ctgtgctcctgcctgtcccgtcacgtgccagcaccctgagccgctggcctgccccgtgcagtgtgtggagggctgtcacg cacactgccctccggnntcaagactgccctgtgtgtgaggcg

gccgggcagcgcttagcgcccgggaagaaaatcaccttgaacgccggggaccctgagcgctgccagatttgtcactgtga tggtgtcagcctcgcctgtgaagcctgcagggaaccgggaggcctggtggtgccccccacagaaggcccggttgacctca ccaccccgtacgtggaggacacgccggagccgcccctgcacgacttcttctgcagcaaactcctggacctggtcttcctg ctggacggctcctccaagctgtctgaggccgacttcgagacgctgaaggcgttcgtggtgggcgtgatggagcgtctgca catctcccagaagcgcatccgcgtggccgtggtggagtaccatgacggctcccacgcctacatcgcactgcaggaccgga agcggccgtcggagctgcggcgcatcgccgggcaggtgaagtacgcgggcagcgaggtggcctccaccagcgaggtcttg aagtacacgctcttccagattttcagcaaggtggaccggtctgagtcctcccgcgtggccctgctgctcacggccagcca ggagccccccaggctggcccggaacttggtnn nn nnnnnnnnnnnnnnnnnnnnnnnnnnnntcaagatcatcagctacctctgcgacctcgcccccgaagcacctgcccccac gcagcggccgctcacggcccaggtctccgtgacgctggcgcccaggaggagctccgtggtcctggatgtggcgttcctcc tggaagggtcggacagcgtgggcgaggccagcttcaacaggagcacagagttcttggaggaggtgatccggcagatggac gtgggccaggacggtgtccacgtcacggtgctgcagtactcgtacgtggtgaccgtggagcactccttcagagaggcgca gtccaaggaggacgtcctgcggcggctgcgggagatccgctaccggggtggcaaccggaccaacaccgggctggccctgc agtacctgtcggagcacagcttctcggccagccagggggaccgggggcaggcgcccaacctggtctacatggtcacggga agcccggcctcagatgagatcaggcggatgcccggagacatccaggtggtgcccattggagtggggcctcgcgtggacgt gccggagctggagaagatcggctggccccaggcccccatcttcatcaaggacttcgagacgctcccccgagaggctcctg acctggtgctgcagaggtgctgctccagagaggggccgcatctccccacctccgcccccgtctcagactgcagccagccc ctggatgtggtcctcctcctggacggctcctccagctttccagcttcttactttgacgaaatgaagagttttgccaaggc tttcatctcaagagccaatataggccctcagctcacccaggtgtcggtgctccagtacggaagtatcaccaccatcgacg tgccgtggaccgtgcccctggagaaagaccacttactgagcctcgtggaccacatgcagcaggagggcggccccagccaa atcgnn nngagtggcggtgt tccccattgggattggagacgggtatgacttggcccagctgaggatcttggcgggcccaggggccagctccaacgtggtg cagctccagcggatcgaagacctccccaccatggtcgccctgggcaattccttcctccacaagctgtgctctgggttcat cagtgtttgcgtggatgaggacgggaatgagaggaggcccggggacatctggaccttgccggatcggtgccacacagtga cctgcctgccagatggccagaccttgctgaatagtcaccgggtcaactgtgaccagcagctgcggccttcctgccccaac agccagtccccgatcagggtggaagaggcctgtggctgccgctggacctgcccctgtgtctgcactggcagctctactcg gcacatcgtgacctttgatggacagaatttcaagctgactggcaactgctcgtacgtcttatttcacaacaaggagcaag acctggaggtactgctccataatgacgcctgcagcgccagggcaaggcagacctgcatgaagtccatcgaggtgaaacac aacggtgtctcggttgagctccgcagcaacatggaggtggtggtgaacgggaggctggtctctgccccttatgtgggtgg ggacatggaggtcagtgtctacggtgccattatgtacgaggtcagactcaagcatctgggccacatcctcaccttcaccc cacaaaacaacgagttccagctgcagctgagccccaagacctttgcctcaaagatgtacggtctctgtggcatctgtgat gagaatggggccaacgacttcatgctgagggatggcacagtcacctcagactggaagacgctggtccaggagtggacggt gcagaagccggggcagacatgcctgctgggtcccgaggagccatgtcccatctccaggggctcccactgccaggtcctcc tctcggagctgtttgccgagtgccacgcggtccttgccccggccacatttcacgccatctgccagcaggacagttgccac caggagcaagtgtgtgaggccattgcctcttatgcccacctctgtcggaccaagggggtgtgtgtcgactggaggacccc cgatttctgtgctgtaccgtgcccgccatccctggtctacaaccactgtgagcgcggctgcccccgccaatgtgagggca actcgagctcctgtggggaccatccctcagaaggctgcttctgccccctgcaccacgtgctgctggaaggcagctgtgtc cccgaggaggcctgcacccagtgcgtcggtgacgatgggatccggcaccagcttctggaaacgtgggtcccagaccacca gccctgccagatctgcatgtgcctcagcgggcggaaggtcaactgcacgacacagccctgccccacagcccgagcgcctg cgtgcggcccgtgcgaggtggcccgtctccggcagagcacgcagcagtgctgccccgagtacgagtgtgtgtgtgacctg gtgagctgtgacctgcccccggtgcctcactgcgaaggtggcctccagcctacactgaccaaccccggcgagtgcagacc caacttcacctgctcctgcaggaaggaggaatgcccaagggcgtccccgccctcctgtcccccgcaccggacgccggtcc ttcagaagacccagtgctgtgatgagtataagtgtgtttgcaactgtaccagctcggcggtgacctgcccactggggtac ctggcctccaccgtcaccaacgactgcggctgcaccaccaccacctgcctccctgataaggtgtgtgtccaccaaggcac catctaccctgtgggccagttctgggaggagggctgtgacgtgtgcacctgcacggacctggaggatgccgtgatgggcc tgcgtgtggcccagtgctctcagaaggcctgtggagacagctgccggccgggcttcacttacgttctccacgaaggcgag tgctgtggaacgtgcctgccatctgcctgcgaggtggtgaccggctcccctcggggagactcgcagtcctcctggaagag tgttggctctcactgggcctctcccgaggacccctgcctcatctacgagtgtgtccgagtcaaggaggaggtcttcgtgc aacagaggaatgtctcctgcccccaactggacgtccccacttgcccctcgggcttccagctgagctgtcagacctcggac tgttgccccacctgtcactgcgagcctgtgcaggcctgtgtgctcaacggcaccatcatcgggcccgggaggagtctgat ggttgacctgtgtaccacctgcctctgcaccgtccaggcgggggtcatctccggattcaagctggagtgcaggcagacca cctgcgaggactgcccgctgnn nnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnncgcgacgagactctccaggatggctgtgacagtcacttctg cagggtcaacaagagaggcgagctcatctgggagaagagggtcacgagctgcccgcccttcgatggacacaagtgtctgg ccgagggagggaaaatcatgaaaattccaggcacatgctgtgacacatgtgaggagcccgagtgcaaggacatcactgcc aagctgcagtacgtcaaagtgggagactgtctgtctgaagagcaagtggacattcactactgcgagggaagatgcaccag caaagccctgtactccatcgacacagaggacgtgcaggaccagtgcgcctgctgcgcgcccacacgcacagagcccatgc aggtgcccctgcgctgcgccaacggctccgtcatccaccacgtggtcctcaacgccatgcagtgcaagtgctcccccagg aagtgccgccagtgacacccgtcaccacccgacttggcaactcctgcttggcgcatgcgtgctccgctctcgggtcttgc tcgggcccacaataaaggtcaatcctgtcatcttgt

Supplemental Figure Legends

FIG. S1. **Phylogenetic tree of vertebrates with sequenced VWF genes.** A taxonomic tree was constructed with the "Common Tree" tool and the lineages in the NCBI taxonomy database for the species indicated (http://www.ncbi.nlm.nih.gov/guide/taxonomy/, accessed November 1, 2010).

FIG. S2*.* **Sequence pair distances for VWF.** The protein sequences of VWF were aligned with MegAlign (version 8.1.5, DNASTAR, Inc., Madison, WI) using the Clustal W method and Gonnet protein weight matrices. For the aligned sequences, the graph displays percent identity directly above the diagonal, and divergence by comparing sequence pairs in relation to the reconstructed phylogeny below the diagonal.

FIG. S3. **Alignment of VWF domains D1D2D'D3**. Protein sequences of VWF corresponding to the D1D2D'D3 domains of human VWF (residues 23-1227) were aligned with MegAlign using the Clustal W method and Gonnet protein weight matrices. Residues that match human VWF exactly are boxed and shaded *yellow*. His residues are shaded *green* if present in human VWF and *blue* if not.

FIG. S4. **Alignment of D1D2D'D3 domains for human VWF and vertebrate gel-forming mucins**. Protein sequences of human VWF corresponding to the D1D2D'D3 domains (residues 23-1227) were aligned to the indicated representative gel-forming mucins with MegAlign using the Clustal W method and Gonnet protein weight matrices. Residues that match human VWF exactly are boxed and shaded *yellow*. His residues are shaded *green* if present in human VWF and *blue* if not.

FIG. S5. **Expression of separate VWF D1D2 and D'D3 constructs with His mutations**. BHK-fur4 cells were transiently transfected with wild type (WT) or mutant D1D2 and D'D3 constructs having the indicated amino acid substitutions. Conditioned medium (A) and cell lysates (B) were analyzed by SDS-PAGE and Western blotting with HRP-conjugated anti-VWF. Positions corresponding to monomeric D'D3, D1D2, and dimeric $(D'D3)_2$ are indicated at the left.

Supplemental References

- 1. Sadler, J. E., Shelton-Inloes, B. B., Sorace, J. M., Harlan, J. M., Titani, K., and Davie, E. W. (1985) *Proc. Natl. Acad. Sci. USA* **82**, 6394-6398
- 2. Shelton-Inloes, B. B., Broze, G. J., Jr., Miletich, J. P., and Sadler, J. E. (1987) *Biochem. Biophys. Res. Commun.* **144**, 657-665
- 3. Bonthron, D. T., Orr, E. C., Mitsock, L. M., Ginsburg, D., Handin, R. I., and Orkin, S. H. (1986) *Nucleic Acids Res.* **14**, 7125-7127
- 4. Mancuso, D. J., Tuley, E. A., Westfield, L. A., Worrall, N. K., Shelton-Inloes, B. B., Sorace, J. M., Alevy, Y. G., and Sadler, J. E. (1989) *J. Biol. Chem.* **264**, 19514-19527
- 5. Mancuso, D. J., Tuley, E. A., Westfield, L. A., Lester-Mancuso, T. L., Le Beau, M. M., Sorace, J. M., and Sadler, J. E. (1991) *Biochemistry* **30**, 253-269
- 6. Dogan, R. I., Getoor, L., Wilbur, W. J., and Mount, S. M. (2007) *Nucleic Acids Res* **35**, W285-291

Dang et al Figure S2

All Match Human 28/30 Match Human
Majority

VWF_H_sapiens VWF_P_troglodytes VWF_G_gorilla
VWF_P_abelii VWF_M_mulatta VWF_C_jacchus VWF_L_africana VWF_A_melanoleuca VWF_P_vampyrus
VWF_B_taurus VWF_T_truncatus
VWF_S_scrofa VWF_D_novemcinctus VWF_O_cuniculus VWF_C_porcellus
VWF_M_musculus VWF_R_norvegicus VWF_M_domesticus
VWF_G_gallus VWF_T_guttata VWF_A_carolinensis VWF_X_tropicalis VWF_G_aculeatus
VWF_O_latipes VWF_T_nigroviridis VWF_T_rubripes

All Match Human VWF Majority Match Human VWF
Majority VWF_H_sapiens

MUC2_H_sapiens
MUC2_C_lupus MUC2_E_caballus MUC2_M_musculus MUC2_R_norvegicus MUC2_G_gallus MUC5AC_H_sapiens MUC5AC_C_lupus MUC5AC_E_caballus MUC5AC_M_musculus MUC5AC_R_norvegicus MUC5B_H_sapiens MUC5B_C_lupus MUC5B_M_musculus MUC5B_R_norvegicus MUC6_H_sapiens
MUC6_C_lupus MUC6_M_musculus MUC6_R_norvegicus MUC6_G_gallus MUC19_H_sapiens MUC19_C_lupus MUC19_B_taurus MUC19_S_scrofa MUC19_M_musculus MUC19_R_norvegicus FIMB1_X_laevis

Dang et al Figure S5