

Protein sequence alignment of COL2A1 protein. COL2A1 protein show high conservation of primary structure among distant

Supplementary material

<https://www.uniprot.org>

CLUSTAL O(1.2.4) multiple sequence alignment

```
SP|P02458|CO2A1_HUMAN  MIRLGAPQTLVLL---TLLVAAVLRCQG-QDVQEAGSCVQDGQRYNDKDVWKPEPCRICV 56
SP|P28481|CO2A1_MOUSE MIRLGAPQSLVLL---TLLIAAVLRCQG-QDAQEAGSCLQNGQRYKDKDVWKPSSCRICV 56
SP|Q91717|CO2A1_XENLA  MFSFVDSRTLVLFAATQVILLAVVRCQDEEDVLDTGSCVQHGRYSDKDVWKPEPCQICV 60
TR|Q2LDA1|Q2LDA1_DANRE MFRL LDSRTL LLLVATHSVLLSLVRCQEQEDDQEEFGGCVQDGQQYADRAVWKPEPCRVCV 60
      *:  :   :*:*:   : : :*:**  :*  : *.*:*.**: * * : ***.  *::**

SP|P02458|CO2A1_HUMAN  CDTGTVLCDDIICEDVKDCLSP EIPFGECCPICPTDLAT-ASGQPGPKGQKGE PGDIKDI 115
SP|P28481|CO2A1_MOUSE  CDTGNVLCDDIICED-PDCLNPEI PFGECCPICPADLAT-ASGKLGPKGQKGE PGDIRDI 114
SP|Q91717|CO2A1_XENLA  CDTGTVLCDDIICEESKDCPN AEI PFGECCPICPTEQSSTSSGQV LKGQKGE PGDIKDV 120
TR|Q2LDA1|Q2LDA1_DANRE CDSGTVLCDEVICEDLND CANPI I SPGECCPICPADTDD-PIGSLGAKGQKGE PGDITDV 119
      **:*.****: :***:  ** .  *  *****: :      * .  *****  *:

SP|P02458|CO2A1_HUMAN  VGPKGPPGPQGPAGEQGP RGRDGDKGEKGAPGPRGRDGEPTPGNPGPPGPPGPPGGL 175
SP|P28481|CO2A1_MOUSE  IGPRGPPGPQGPAGEQGP RGRDGDKGEKGAPGPRGRDGEPTPGNPGPAGPPGPPGPPGGL 174
SP|Q91717|CO2A1_XENLA  LGPRGPPGPQGPSGEQGS RGERGDKGEKGAPGPRGRDGEPTPGNPGPV--GPPGPPGGL 177
TR|Q2LDA1|Q2LDA1_DANRE VGP RGPAGPMGPPGEQGT RGERGAKGEKGS PGPRGRDGEPTPGNPGPPGPPGPPGPPGGL 179
      :**:* ** ** ** ** ** ** ** ** ** ** ** ** ** ** ** **^*****:***** ** *****

SP|P02458|CO2A1_HUMAN  G-GNFAAQMAGGFDEKAGGAQLGVMQGPMPGMPGRGPPG PAGAPGPQGFQGNPGEPEPG 234
SP|P28481|CO2A1_MOUSE  SAGNFAAQMAGGYDEKAGGAQMGVMQGPMPGMPGRGPPG PAGAPGPQGFQGNPGEPEPG 234
SP|Q91717|CO2A1_XENLA  G-GNFAAQMTGGFDEKAGGAQMGVMQGPMPGMPGRGPPG PGTGAPGPQGFQGNPGEPEPG 236
TR|Q2LDA1|Q2LDA1_DANRE G-GNFAAQMAGGFDEKAGGAQMGVMQGPMPGMPGRGPPG PPSGAPGPQGFQGNPGETGEPG 238
      . *****: ** :*****:*****:*****:*****:***** *****

SP|P02458|CO2A1_HUMAN  VSGPMGPRGPPGPPGKPGDDGEAGKPGKAGERGPPGPQARGFPPTPGLPGVKGHRGYPG 294
SP|P28481|CO2A1_MOUSE  VSGPMGPRGPPGPPGKPGDDGEAGKPGKSGERGLPGPQARGFPPTPGLPGVKGHRGYPG 294
SP|Q91717|CO2A1_XENLA  AGGPMGPRGPPGPPSGKPGDDGEAGKPGKSGERGPPGPQARGFPPTPGLPGVKGHRGYPG 296
TR|Q2LDA1|Q2LDA1_DANRE PAGALGPRGPPGPPGKPGSDGEAGKPGKAGERGPPGPQARGFPPTPGLPGIKGHRGHPG 298
```

. * :***** ****.*****:**** *****:*****:****:****

SP	P02458	CO2A1_HUMAN	LDGAKGEAGAPGVKGESGSPGENSGPMPGPRGLPGERGRTGPAGAAGARGNDGQPPGAG	354
SP	P28481	CO2A1_MOUSE	LDGAKGEAGAPGVKGESGSPGENSGPMPGPRGLPGERGRTGPAGAAGARGNDGQPPGAG	354
SP	Q91717	CO2A1_XENLA	LDGAKGEAGAAGAKGEGGATGEAGSPGMPGPRGLPGERGRPGSSGAAGARGNDGLPGPAG	356
TR	Q2LDA1	Q2LDA1_DANRE	LDGAKGEAGAAGAKGEGSNGESGAPGMPGPRGLPGERGRPGATGAAGARGNDGLPGPAG	358

***** * .***. *: ** * :***** * .***** *****

SP	P02458	CO2A1_HUMAN	PPGPVGPAGGPGFPGAPGAKGEAGPTGARGPEGAQGPGEPTPGSPGPAGASGNPGTDG	414
SP	P28481	CO2A1_MOUSE	PPGPVGPAGGPGFPGAPGAKGEAGPTGARGPEGAQGSRGEPGNPGSPGPAGASGNPGTDG	414
SP	Q91717	CO2A1_XENLA	PPGPVGPAGAPGFPAGPSKGEAGPTGARGPEGAQGPRESGTPGSPGPAGASGNPGTDG	416
TR	Q2LDA1	Q2LDA1_DANRE	PPGPVGPAGAPGFPSPGSKGEAGPTGARGPEGAQGPGEAGTPGSPGPAGASGNPGTDG	418

***** .*****:***:***** ***** * .***** *****

SP	P02458	CO2A1_HUMAN	IPGAKGSAGAPGIAGAPGFPGRGPPGPGQATGPLGPKGQTGEPGIAGFKGEQGPKEGEPG	474
SP	P28481	CO2A1_MOUSE	IPGAKGSAGAPGIAGAPGFPGRGPPGPGQATGPLGPKGQAGEPGIAGFKGDQGPKEGEPG	474
SP	Q91717	CO2A1_XENLA	IPGAKGSSGGPGIAGAPGFPGRGPPGPGQATGPLGPKGQTDGDPGVAGFKGEQGPKEIG	476
TR	Q2LDA1	Q2LDA1_DANRE	IPGAKGSAGASGIAGAPGFPGRGPPGPGQATGPLGPKGQSGDPIPGFKGEAGPKGERG	478

*****:*. *****:***:****: ***** *

SP	P02458	CO2A1_HUMAN	PAGPQGAPGPAGEEGKRGARGEPPGVPIGPPGERGAPGNRGFPQDGLAGPKGAPGERG	534
SP	P28481	CO2A1_MOUSE	PAGPQGAPGPAGEEGKRGARGEPPGAGPIGPPGERGAPGNRGFPQDGLAGPKGAPGERG	534
SP	Q91717	CO2A1_XENLA	SAGPQGAPGPAGEEGKRGARGEPPAAGPNPPGERGAPGNRGFPQDGLAGPKGAPGERG	536
TR	Q2LDA1	Q2LDA1_DANRE	VLGPQGPSPGSEEGKRGPRGEPGSAGPLGPPGERGAPGNRGFPQDGLAGKAGAPDRG	538

**** * :***** ***** . ** ***** *****:***

SP	P02458	CO2A1_HUMAN	PSGLAGPKGANGDPGRGPEPLPGARLTGTRPGDAGPQKVGPSGAPGEDGRPGPPGPGQ	594
SP	P28481	CO2A1_MOUSE	PSGLTGPKANGDPGRGPEPLPGARLTGTRPGDAGPQKVGPSGAPGEDGRPGPPGPGQ	594
SP	Q91717	CO2A1_XENLA	VPGLGGPKGNGDPGRGPEPLPGARLTGTRPGDAGPQKVGPSGASGEDGRPGPPGPGQ	596
TR	Q2LDA1	Q2LDA1_DANRE	VPGLSGPKGTDGDPGRGPEPLPGARLTGTRPGDAGAQQKVGATGAPGEDGRPGPPGPLG	598

** **** . ***** ***** :** ***** *

SP	P02458	CO2A1_HUMAN	ARGQPGVMGFPGPKGANGEPGKAGEKGLPGAPGLRGLPGKDGETGAAGPPGPAGPAGERG	654
SP	P28481	CO2A1_MOUSE	ARGQPGVMGFPGPKGANGEPGKAGEKGLAGAPGLRGLPGKDGETGAAGPPGPSGAPAGERG	654
SP	Q91717	CO2A1_XENLA	ARGQPGVMGFPGPKGANGEPGKAGEKGLVAPGLRGLPGKDGETGSQGNPAGPAGERG	656
TR	Q2LDA1	Q2LDA1_DANRE	ARGQPGVMGFPGPKGANGEPGKPEKGLVGRGTGLRGLPGKDGETGPSGPPGPVAVGERG	658

***** ***** * ***** ** ** * .****

SP	P02458	CO2A1_HUMAN	EQGAPGPSGFQGLPGPPGPPGEGGKPGDQGVPEAGAPLVGPRGERGFPGERGSPGAQG	714
SP	P28481	CO2A1_MOUSE	EQGAPGPSGFQGLPGPPGPPGEGGKQGDQGIPEAGAPLVGPRGERGFPGERGSPGAQG	714
SP	Q91717	CO2A1_XENLA	EQGPPGPSGFQGLPGPPGSPGEGGKPGDQGVPEAGAPLVGPRGERGFPGERGSSGPQG	716

TR	Q2LDA1	Q2LDA1_DANRE	EQQQPGPSGFQGLPGPTGAPGEPGKPGDQGVPGEGGAAGPTGPRGERGFPGERGGAGPQG *** ***** * ** ** *****:***.*** * .*****. * **	718
SP	P02458	CO2A1_HUMAN	LQGPRGLPGTPTDGPKGASGPAGPPGAQGGPGLQGMPPERGAAGIAGPKGDRGDVGEKG	774
SP	P28481	CO2A1_MOUSE	LQGPRGLPGTPTDGPKGAAAGPDGPPGAQGGPGLQGMPPERGAAGIAGPKGDRGDVGEKG	774
SP	Q91717	CO2A1_XENLA	LQGPRGLPGTPTDGPKGASGPSGPNQAQGGPGLQGMPPERGAAGISGPKGDRGDTGEKG	776
TR	Q2LDA1	Q2LDA1_DANRE	LQGPRGLPGTPTDGPKGAIGPAGAAGAQGGPGLQGMPPERGAAGISGAKGDRGDSGEKG ***** ** * *****.***:***** **	778
SP	P02458	CO2A1_HUMAN	PEGAPGKDGGRLTGPPIGPPGAGANGEKGEVGGPPGAGSAGARGAPGERGETGPPGPAG	834
SP	P28481	CO2A1_MOUSE	PEGAPGKDGGRLTGPPIGPPGAGANGEKGEVGGPPGSGTGARGAPGERGETGPPGPAG	834
SP	Q91717	CO2A1_XENLA	PEGASGKDGSRGLTGPPIGPPGAGPNGEKGESGSPGPIVGARGAPDRGENGPPGPAG	836
TR	Q2LDA1	Q2LDA1_DANRE	PEGAPGKDGSRGLTGPPIGPPGSPNGAKGETGPIGSIGAPGARGAPDRGEIGAPGPAG **** * .*****.*** ** ** * * *****.*** * ****	838
SP	P02458	CO2A1_HUMAN	FAGPPGADGQPGAKGEQGEAGQKGDAGAPGQGPSGAPGPPQPTGVTGPKGARGAQGGP	894
SP	P28481	CO2A1_MOUSE	FAGPPGADGQPGAKGDQGEAGQKGDAGAPGQGPSGAPGPPQPTGVTGPKGARGAQGGP	894
SP	Q91717	CO2A1_XENLA	FAGPPGADGQSGGLKGDQGESGQKGDAGAPGQGPSGAPGPPQPTGVFGPKGARGAQGP	896
TR	Q2LDA1	Q2LDA1_DANRE	FAGPPGADGQPGNKGEQGESGQKGDAGAPGQGPSGAPGPPVPTGVTGPKGARGAQGAP ***** * **:***:*****:***** ***** ***** *	898
SP	P02458	CO2A1_HUMAN	ATGFPGAAGRVPGGPNNGPPGPPGSGKDGPKGARGDSGPPGRAGEPGLQGPAGPPG	954
SP	P28481	CO2A1_MOUSE	ATGFPGAAGRVPGGANGNPPGPPGPKDGPKGVRGDSGPPGRAGDPGLQGPAGAPG	954
SP	Q91717	CO2A1_XENLA	ATGFPGAAGRVPGPNGNPPGPPGSGKDGPKGVRGDAGPPGRAGDPGLQGAAGAPG	956
TR	Q2LDA1	Q2LDA1_DANRE	ATGFPGAAGRVPGGPNNGPAAAGPAGPSGKDGPKGVRGDAGPPGRAGDAGLRGPPGAP ***** ** ***** ** * **:***:***.***:*****: **:* **	958
SP	P02458	CO2A1_HUMAN	EKGEPGDDGPGSAGPPGPPQGLAGQRGIVGLPGQRGERGFPLPGPSGEPGKQGAPGASG	1014
SP	P28481	CO2A1_MOUSE	EKGEPGDDGPGSLDGPPGPPQGLAGQRGIVGLPGQRGERGFPLPGPSGEPGKQGAPGASG	1014
SP	Q91717	CO2A1_XENLA	EKGEPGEDGPGSPDPPGPPQGLSQRGIVGLPGQRGERGFPLPGPSGEPGKQGGPSSG	1016
TR	Q2LDA1	Q2LDA1_DANRE	EKGAGEDGPPGPDGPGSAGPAGLAGQRGIVGLPGQRGERGFPLPGPSGEPGKQGGPSSG **** *:*** * **:*** ** **:*****:*****:*****.***.***	1018
SP	P02458	CO2A1_HUMAN	DRGPPGVPVGGPGLTGPAGEPREGSPGADGPPGRDGAAGVKGDRGETGAVGAPGAPGPPG	1074
SP	P28481	CO2A1_MOUSE	DRGPPGVPVGGPGLTGPAGEPREGSPGADGPPGRDGAAGVKGDRGETGALGAPGAPGPPG	1074
SP	Q91717	CO2A1_XENLA	DRGPPGVPVGGPGLTGPAGEPREGSPGADGPPGRDGAAGVKGDRGETGALGAPGAPGAPG	1076
TR	Q2LDA1	Q2LDA1_DANRE	DRGPPGVPVGGPGLTGPAGEPREGSPGADGPPGRDGAAGVKGERGNTGPIGAPGAPGAPG *****:*** *****.***:*****:***:***:***:***:***** **	1078
SP	P02458	CO2A1_HUMAN	SPGPAGPTGKQDRGEAGAQQGPMGSPGAGARGIQGPQPRGDKGEAGEPGERGLKGHRG	1134
SP	P28481	CO2A1_MOUSE	SPGPAGPTGKQDRGEAGAQQGPMGSPGAGARGIAGPQPRGDKGESGEQGERGLKGHRG	1134

SP	Q91717	CO2A1_XENLA	APGSVGP	TGKQ	DRGESG	PQGPLG	PSGPAG	ARGLAG	PQGPRG	DKGEAGE	EAGERG	QK	GHRG	1136				
TR	Q2LDA1	Q2LDA1_DANRE	APGSVGP	IGKQ	DRGENG	PQGPAG	PPGPAG	ARMVGP	QGPRG	DKGEAGE	EAGERG	QK	GHRG	1138				
			: ** . ** ***** * ** * ** ***** : ***** : ** * ** * ** *															
SP	P02458	CO2A1_HUMAN	FTGLQGL	PGPPG	PSGDQ	GASGP	PAGPSG	PRGPPG	PV	PSGK	DGANGI	PGPI	GPPG	PRGRSG	1194			
SP	P28481	CO2A1_MOUSE	FTGLQGL	PGPPG	PSGDQ	GASGP	PAGPSG	PRGPPG	PV	PSGK	DGSNGI	PGPI	GPPG	PRGRSG	1194			
SP	Q91717	CO2A1_XENLA	FTGLQGL	PGPPG	SAGDQ	GATGP	PAGPAG	PRGPPG	PV	PSGK	DGSNGI	SGPI	GPPG	PRGRSG	1196			
TR	Q2LDA1	Q2LDA1_DANRE	FTGLQGL	PGPPG	SPGDQ	GAAGP	AGPSG	AKGPSG	PV	PSGK	DGSNGI	QPGPI	GPPG	PRGRSG	1198			
			***** ***** : ***** : ** * ** ***** : ***** : ** * ** *****															
SP	P02458	CO2A1_HUMAN	ETGPAGP	PGNPG	PPGPP	PGPPG	IDMSA	FAGLGP	PREKGP	DP	LQYMRAD	QAAGGL	RQ	HDAE	1254			
SP	P28481	CO2A1_MOUSE	ETGPVGP	PGSPG	PPGPP	PGPPG	IDMSA	FAGLQ	REKGP	DP	QYMRAD	EADSTL	RQ	HDVE	1254			
SP	Q91717	CO2A1_XENLA	ETGPGSP	PGQPG	PPGPP	PGPPG	IDMSA	FAGLSQ	PEKGP	DP	MYMRAD	QASNSL	- - -	PVD	1253			
TR	Q2LDA1	Q2LDA1_DANRE	ESGPVGP	PGNPG	PPGPP	PGPPG	IDMSA	FAGLSQ	PEKGP	DP	LYMRAD	EASSSL	RQ	HDVE	1258			
			* : ** * ** . ***** . ***** . ***** : ***** : * . * . :															
SP	P02458	CO2A1_HUMAN	VDATLKS	LNNQ	IESIR	SPESR	RKNPART	CRDLK	LCHPEW	KSGDY	WIDPNQ	GCTL	DAMK	VF	1314			
SP	P28481	CO2A1_MOUSE	VDATLKS	LNNQ	IESIR	SPDGS	RKNPART	CQDLK	LCHPEW	KSGDY	WIDPNQ	GCTL	DAMK	VF	1314			
SP	Q91717	CO2A1_XENLA	VEATLKS	LNNQ	IENIR	SPDGT	KKNPART	CRDLK	LCHPEW	KSGDY	WIDPNQ	GCTV	DAIK	VF	1313			
TR	Q2LDA1	Q2LDA1_DANRE	VDATLKS	INGQ	IEDIR	SPDGS	RKNPAR	SCRDLK	LCHPEW	KSGDY	WVDP	NL	GSAADAI	KVF	1318			
			* : ***** : * . ** . ***** : * : ***** : * : ***** : ***** : ** * . : ** : ** *															
SP	P02458	CO2A1_HUMAN	CNMETG	ETCVY	PNPAN	VPKKN	NWSSK	SKEK	KHIWF	GETING	GGFH	FSYGD	DN	LAPNTAN	VQ	1374		
SP	P28481	CO2A1_MOUSE	CNMETG	ETCVY	PNPAT	VPRKN	NWSSK	SKEK	KHIWF	GETING	MGGF	HFSYGD	GN	LAPNTAN	VQ	1374		
SP	Q91717	CO2A1_XENLA	CDMETG	ETCVY	PNPSK	IPKKN	NWSSA	KGK	KHIWF	GETING	GGFQ	FSYGD	DSSA	PNTANI	Q	1373		
TR	Q2LDA1	Q2LDA1_DANRE	CNMETG	ETCVK	PSTPK	IPRKN	NWTSK	SKAQ	KHVWF	GESM	GGF	HFSYAD	GSQ	TPSTTTI	Q	1378		
			* : ***** * . . : * : ***** : * . * : ** : ***** : ***** : ** * . * . . : * . * . : *															
SP	P02458	CO2A1_HUMAN	MTFLRLL	STEGS	QNI	TYHCK	NSIAYL	DEAAGN	LKKALLI	QGSND	VEIRA	EAGNSR	F	TYTAL	1434			
SP	P28481	CO2A1_MOUSE	MTFLRLL	STEGS	QNI	TYHCK	NSIAYL	DEAAGN	LKKALLI	QGSND	VEMRA	EAGNSR	F	TYTAL	1434			
SP	Q91717	CO2A1_XENLA	MTFLRLL	STDAS	QNI	TYHCK	NSIAFM	DEASGN	LKKAVLL	QGSND	VEIRA	EAGNSR	F	TYNAL	1433			
TR	Q2LDA1	Q2LDA1_DANRE	LNFLRLL	STEATQ	TIT	TYHCK	NSVAYM	DQATGN	LKKAILL	QGSND	VEIRA	EAGNSR	F	TYGVL	1438			
			. : ***** : . : * . ***** : * : * : ***** : * : ***** : ***** . * . *															
SP	P02458	CO2A1_HUMAN	KDGCTK	H	TGK	WGKT	VIEYRS	QKTSR	LPIIDI	APMDI	GGPE	QEF	GVDI	GP	VCF	L	1487	
SP	P28481	CO2A1_MOUSE	KDGCTK	H	TGK	WGKT	VIEYRS	QKTSR	LPIIDI	APMDI	GGAE	QEF	GVDI	GP	VCF	L	1487	
SP	Q91717	CO2A1_XENLA	EDGCKK	H	TGK	W	K	TVIEYR	TQKTSR	LPIVDI	APMDI	GGAD	QEF	GVDI	GP	VCF	L	1486
TR	Q2LDA1	Q2LDA1_DANRE	EDGCKK	H	TGQ	WAKT	VIEYK	TQKTSR	LPIMDI	APMDI	GGAD	QEF	GVDI	GA	VCF	L	1491	
			: ** . ***** : * . ***** : : ***** : ***** : ***** : ***** * ** *															

Protein sequence alignment of COL2A1 protein. COL2A1 proteins show high conservation of the primary structure among distant species. The figure shows CLUSTAL OMEGA (1.2.4) multiple sequence alignment of the regions of COL2A1 protein of different organisms: Human, *Mus musculus*, *Xenopus laevis*, and *Danio rerio* [Uniprot identifiers: P02458, P28481, Q91717, Q2LDA, respectively]. Highlighted amino acids show the mutations found in Perthes patients. c.638 G > A (p.Gly213Asp) (exon 9); c.2014 G > T (p.Gly672Cys) (exon 31) (NM_001844.5) [50]; c.1888 G > A (p.Gly630Ser) (exon 29) (NM_033150.3) [51]; c.3665 G > A (p.Gly1170Ser) (exon 50) (NM_033150.3) [48,49].

48. Su, P.; Li, R.; Liu, S.; Zhou, Y.; Wang, X.; Patil, N.; Mow, C.S.; Mason, J.C.; Huang, D.; Wang, Y. Age at onset-dependent presentations of premature hip osteoarthritis, avascular necrosis of the femoral head, or Legg-Calvé-Perthes disease in a single family, consequent upon a p.Gly1170Ser mutation of COL2A1. *Arthritis Rheum.* **2008**, *58*, 1701–1706.

49. Miyamoto, Y.; Matsuda, T.; Kitoh, H.; Haga, N.; Ohashi, H.; Nishimura, G.; Ikegawa, S. A recurrent mutation in type II collagen gene causes Legg-Calvé-Perthes disease in a Japanese family. *Hum. Genet.* **2007**, *121*, 625–629. <https://doi.org/10.1007/s00439-007-0354-y>.

50. Kannu, P.; Irving, M.; Aftimos, S.; Savarirayan, R. Two Novel COL2A1 Mutations Associated with a Legg-Calvé-Perthes Disease-like Presentation. *Clin. Orthop. Relat. Res.* **2011**, *469*, 1785–1790. <https://doi.org/10.1007/s11999-011-1850-x>.

51. Li, N.; Yu, J.; Cao, X.; Wu, Q.-Y.; Li, W.-W.; Li, T.-F.; Zhang, C.; Cui, Y.-X.; Li, X.-J.; Yin, Z.-M.; et al. A novel p. Gly630Ser mutation of COL2A1 in a Chinese family with presentations of Legg-Calvé-Perthes disease or avascular necrosis of the femoral head. *PLoS ONE* **2014**, *9*, e100505.