

# Annotation of a Contiguous Segment of *Quercus Lobata* DNA

---

Devin Parks

Mentor: Prof. Matteo Pellegrini

# Annotation Goal: Complete Reference Genome

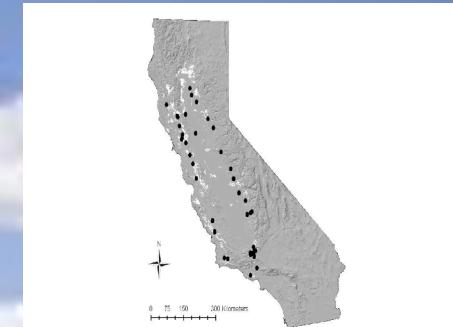
- Genome has been fully sequenced, now it must be annotated to create a reference genome.



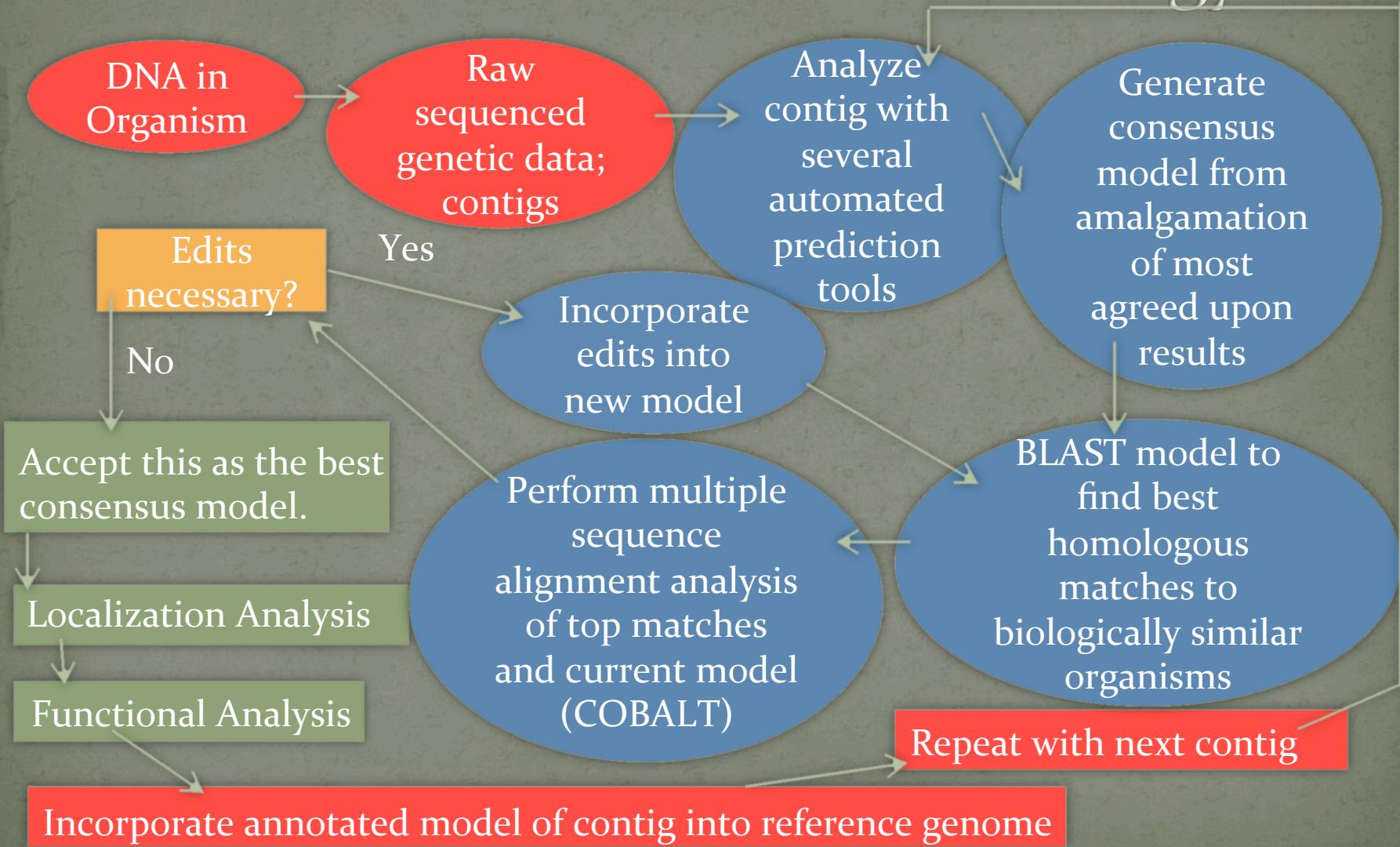
- Automated methods are not currently accurate enough; manual annotation is necessary.
- Essential tool for creation of the most efficient and accurate species-specific conservation plan.

# Quercus Lobata: Common Valley Oak

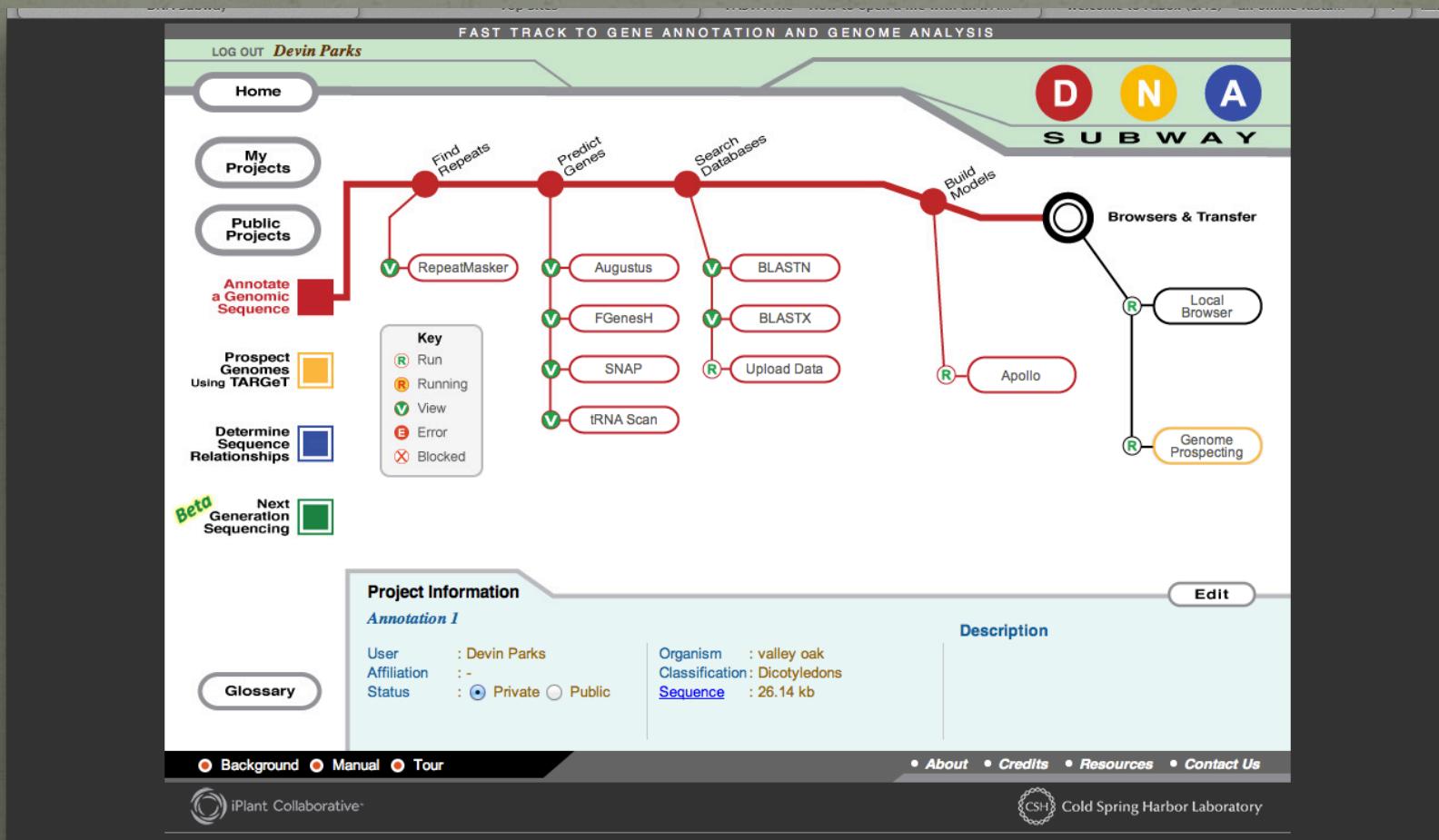
- Common Valley Oak is a foundation species found throughout California.
- *Quercus lobata* exhibits genetic variation that fluctuates respective to geographic location across the state.
- 95% riparian valley oak forest and valley oak woodland have been removed from San Joaquin valley.
- Climate change threatens with increased temperatures and extended droughts



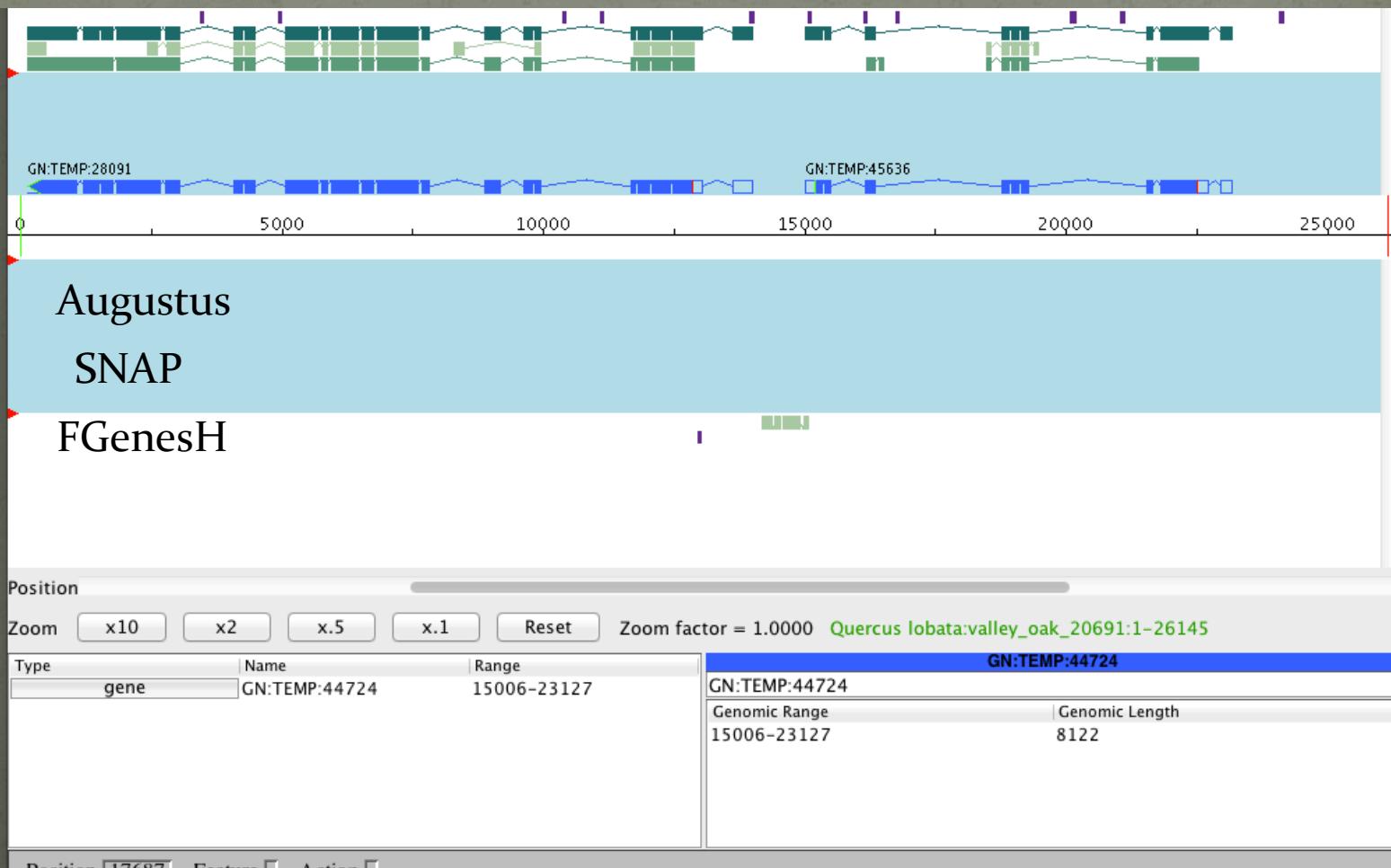
# Manual Annotation: Overall Strategy



# DNA Subway



# Apollo



# Augustus Prediction Fragments

## BLAST Results

### Fragment #1

**Sequences producing significant alignments:**

Select: All None Selected: 0

Alignments Download GenPept Graphics Distance tree of results Multiple alignment

	Description	Max score	Total score	Query cover	E value	Ident	Accession
<input type="checkbox"/>	PREDICTED: putative vacuolar protein sorting-associated protein 13B-like isoform X3 [Citrus sinensis]	2263	2263	99%	0.0	54%	XP_006492901.1
<input type="checkbox"/>	hypothetical protein CICLE_v100108862mg, partial [Citrus clementina] >gb ESR43047.1  hypothetical	2254	2254	99%	0.0	54%	XP_006429807.1
<input type="checkbox"/>	Vacuolar protein sorting-associated protein 13A, putative [Theobroma cacao] >gb EOX92839.1  Vac	2251	2251	99%	0.0	54%	XP_007048682.1
<input type="checkbox"/>	PREDICTED: uncharacterized protein LOC100245550 [Vitis vinifera]	2246	2246	96%	0.0	55%	XP_002275536.2
<input type="checkbox"/>	PREDICTED: uncharacterized protein LOC101298156 [Fragaria vesca subsp. vesca]	2223	2223	99%	0.0	54%	XP_004305785.1
<input type="checkbox"/>	hypothetical protein POPTR_0001s45980g [Populus trichocarpa] >gb ERP67310.1  hypothetical pro	2208	2208	99%	0.0	53%	XP_006370741.1
<input type="checkbox"/>	Putative vacuolar protein sorting-associated protein 13C [Morus notabilis]	2188	2188	99%	0.0	54%	EXB26144.1
<input type="checkbox"/>	hypothetical protein PRUPE_ppa000018mg [Prunus persica] >gb EMJ18863.1  hypothetical protein	2133	2438	99%	0.0	60%	XP_007217664.1
<input type="checkbox"/>	vacuolar protein sorting-associated protein, putative [Ricinus communis] >gb EEF43780.1  vacuolar	2125	2125	99%	0.0	51%	XP_002518393.1
<input type="checkbox"/>	hypothetical protein PHAVU_002G3286000g, partial [Phaseolus vulgaris] >gb ESW32513.1  hypoth	1892	1892	99%	0.0	48%	XP_007160519.1
<input type="checkbox"/>	PREDICTED: LOW QUALITY PROTEIN: uncharacterized protein LOC101204937 [Cucumis sativus]	1871	1871	99%	0.0	48%	XP_004139161.1
<input type="checkbox"/>	PREDICTED: LOW QUALITY PROTEIN: uncharacterized LOC101204937 [Cucumis sativus]	1839	1839	99%	0.0	47%	XP_004157156.1
<input type="checkbox"/>	hypothetical protein PHAVU_002G3286000g [Phaseolus vulgaris] >gb ESW32514.1  hypothetical pr	1815	1815	93%	0.0	49%	XP_007160520.1

### Fragment #2

Alignments Download GenPept Graphics Distance tree of results Multiple alignment

	Description	Max score	Total score	Query cover	E value	Ident	Accession
<input type="checkbox"/>	hypothetical protein PRUPE_ppa000018mg [Prunus persica] >gb EMJ18863.1  hypothetical protein	669	669	87%	0.0	80%	XP_007217664.1
<input type="checkbox"/>	PREDICTED: uncharacterized protein LOC101298156 [Fragaria vesca subsp. vesca]	661	661	87%	0.0	80%	XP_004305785.1
<input type="checkbox"/>	vacuolar protein sorting-associated protein, putative [Ricinus communis] >gb EEF43780.1  vacuolar	661	661	87%	0.0	78%	XP_002518393.1
<input type="checkbox"/>	Putative vacuolar protein sorting-associated protein 13C [Morus notabilis]	660	660	88%	0.0	79%	EXB26144.1
<input type="checkbox"/>	PREDICTED: putative vacuolar protein sorting-associated protein 13B-like isoform X3 [Citrus sinensis]	659	659	88%	0.0	77%	XP_006492901.1
<input type="checkbox"/>	hypothetical protein POPTR_0001s45980g [Populus trichocarpa] >gb ERP67310.1  hypothetical pro	649	649	88%	0.0	78%	XP_006370741.1
<input type="checkbox"/>	Vacuolar protein sorting-associated protein 13A, putative [Theobroma cacao] >gb EOX92839.1  Vac	652	652	88%	0.0	78%	XP_007048682.1
<input type="checkbox"/>	hypothetical protein CICLE_v100108862mg, partial [Citrus clementina] >gb ESR43047.1  hypothetical	641	641	88%	0.0	77%	XP_006429807.1
<input type="checkbox"/>	unnamed protein product [Vitis vinifera]	618	618	88%	0.0	79%	CBI40035.3
<input type="checkbox"/>	PREDICTED: uncharacterized protein LOC100783352 isoform X4 [Glycine max]	619	619	88%	0.0	73%	XP_006594305.1
<input type="checkbox"/>	PREDICTED: uncharacterized protein LOC100783352 isoform X3 [Glycine max]	617	617	88%	0.0	73%	XP_006594304.1
<input type="checkbox"/>	PREDICTED: uncharacterized protein LOC100783352 isoform X1 [Glycine max]	617	617	88%	0.0	73%	XP_006594302.1
<input type="checkbox"/>	C-5 cytosine-specific DNA methylase; Vacuolar protein sorting-associated protein [Medicago truncat	596	596	88%	0.0	70%	ABN08303.1
<input type="checkbox"/>	hypothetical protein PHAVU_002G3286000g [Phaseolus vulgaris] >gb ESW32514.1  hypothetical pr	607	607	88%	0.0	72%	XP_007160520.1
<input type="checkbox"/>	hypothetical protein PHAVU_002G3286000g, partial [Phaseolus vulgaris] >gb ESW32513.1  hypoth	608	608	88%	0.0	72%	XP_007160519.1
<input type="checkbox"/>	hypothetical protein MIMGU_mgv1a0233772mg, partial [Mimulus guttatus]	565	565	87%	0.0	68%	EYU33513.1

# COBALT Data: Block Format

Vacuolar Protein 13B like Isoform X3. Citrus Sinensi

Fragment 1

Fragment 2

Vacuolar Protein 13A Theobromo cacao

Fragment 1

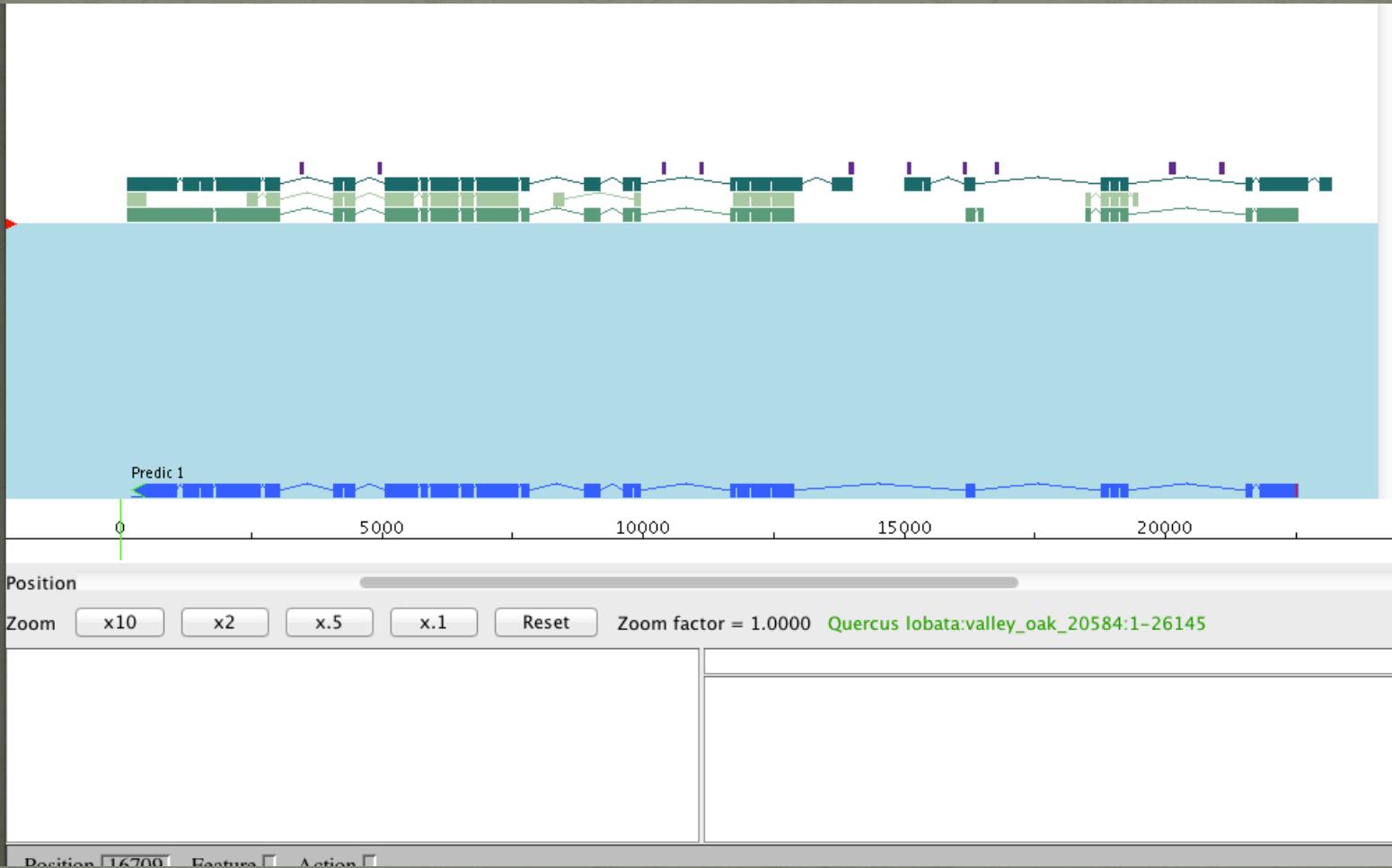
Fragment 2

Uncharacterized Protein LOC101298156 Fragaria vesca

Fragment 1

Fragment 2

# Initial Consensus Model



# COBALT data

9 rows of data correspond to 8 closest homologous matches and our consensus model:

- 1) Putative vacuolar protein sorting-associated protein 13B-like isoform X<sub>3</sub> [Citrus sinensis]
- 2) Vacuolar protein sorting-associated protein 13A putative [Theobroma cacao]
- 3) Hypothetical protein CICLE\_v100108862 [Citrus clementina]
- 4) Uncharacterized protein LOC101298156 [Fragaria vesca subsp. Vesca]
- 5) Uncharacterized protein LOC100245550 [Vitis vinifera]
- 6) Putative vacuolar protein sorting-associated protein 13C [Morus notabilis]
- 7) Hypothetical protein PRUPE\_ppaoooo18mg [Prunus persica]
- 8) Vacuolar protein sorting-associated protein [Ricinus communis]
- 9) Consensus prediction model.

# COBALT data

DNA Subway	2467 residues	COBALT: - Co...	Blast:GN:TEMP...	COBALT: - Co...	Sequence Align...	2564 residues
<input checked="" type="checkbox"/> <a href="#">XP_006429807</a>	551	TFH GSL WGS SLP NLP PI LNL RVR KGN VGS LSS QLE V S I G I Q H V S C V L P P E Y L A I I G Y F S L P D W S P Y	-----	LSEH-NEQ	624	
<input checked="" type="checkbox"/> <a href="#">XP_004305785</a>	1199	NLR FFL WGP SIS N I S P I L N V R V R K E - C G P L S R V E L C F S V Q H V Y C I L P P E Y L A I I G Y F S L P D W S S D	SNQLVTTG-HED	1276		
<input checked="" type="checkbox"/> <a href="#">XP_002275536</a>	1755	NFHEFLWGPSPNLSPILNIRMTKGNAESIGSHSSESLSI	QHVCCILPPEYLAIVIGYFSLPDWGLNANKQPVFGK-HKH	1833		
<input checked="" type="checkbox"/> <a href="#">EXB26144</a>	1011	NFHFIFIWGPCLPNLSPINVRVQKNKHSLSGSHLEVGLSLQH V Y C I L P P E Y L A I I G Y F S L P D W S S D	SNK--TGDGMHES	1088		
<input checked="" type="checkbox"/> <a href="#">XP_007217664</a>	310	NLRQFLWGPSPLSPLSPILNVRVSKEKGCPSSLRIEVSGVQH V Y C I L P P E Y L A I I G Y F S L P D W S S D	SNEQPVIVG-C E Y	388		
<input checked="" type="checkbox"/> <a href="#">XP_002518393</a>	1166	NLKDFI WGP SIS N - P S I L N L R V K K G L G S G V T S Q F E V S I G I Q H V Y C F L P P E Y L A I I G Y F S S S D W S T N L	SMQLVTE N-CDC	1243		
<input checked="" type="checkbox"/> 10001	275	TFR-----	YFALS DWNSN SNDQPVTDG-HEY	299		
<input checked="" type="checkbox"/> <a href="#">XP_006429201</a>	1246	I Y S E N A S I L Y K F E V V D S T L T V P V E K D D N Q L L K V E I Q Q L Y C S F I D K C A S N S V M M D I P P K Y M V P V N K L A E N N D C L N I F G R D	1325			
<input checked="" type="checkbox"/> <a href="#">XP_007048682</a>	1249	M D S Q S E N A I I Y K F E V V L E S T L I L P V E S D H D Q F L K T E I Q Q L Y G S F I D E C A L S D V L K D I P P E Y Y V P E N K V A R T N H C L N I F G R D	1328			
<input checked="" type="checkbox"/> <a href="#">XP_006429807</a>	625	I Y S E N A S I L Y K F E V V D S T L T V P V E K D D N Q L L K V E I Q Q L Y C S F I D K C A S N S V M M D I P P K Y M V P V N K L A E N N D C L N I F G R D	704			
<input checked="" type="checkbox"/> <a href="#">XP_004305785</a>	1277	T E S D N E C S V V Y K I E I L D S L I V P V K S N G S Q F L N L D I Q Q L Y C S F M D K S C S G E V I R D I P P E C L V Q A H E V A D K C S L N V F G R D	1355			
<input checked="" type="checkbox"/> <a href="#">XP_002275536</a>	1834	I N R E P E S D F I L F K L E I V D S T L I L P V K S N G S Q F L N L D I Q Q L Y C S F M D K S C S G E V I R D I P P E C L V Q A H E V A D K C S L N V F G R D	1913			
<input checked="" type="checkbox"/> <a href="#">EXB26144</a>	1089	S D A E N E G S V V Y K F E I L D S V L I L P V E S S E P Q F L K A E F Q Q V V F S F I N S S S P D N A L K G I P C E S L V P E D K L A K R S D C L N I F G R D	1168			
<input checked="" type="checkbox"/> <a href="#">XP_007217664</a>	389	T E T G N E C S L V V Y K I E I L D S T L I L P M K S N E G H F L K T E I Q Q L Y C S F I D S S S L N N V L N D I P P E C W V P A H K L S E R N H C L N L F G R D	468			
<input checked="" type="checkbox"/> <a href="#">XP_002518393</a>	1244	I V T E K G N P V V Y K F E I L D S I L I L P V E R D D H Q F L K A E L Q Q L Y C S I I L N C S P D D V C P I C E C M V P T D K V A K A N D C L N I Y G R D	1323			
<input checked="" type="checkbox"/> 10001	300	I N H A N E D S F I Y K F E I L D S T L I L P V E S N E C Q F L K V E I R Q L Y C S F I H E R I S D D A L K E I P R E -----	358			
<input checked="" type="checkbox"/> <a href="#">XP_006429201</a>	1326	L I L S F V I L K D G G Y G C F L G E Q D P G N -- r N I I L M A P V S A D V W V R I P W E D K P N S E G S I A S T C I M S R I Q N C Q I I V D D C Y A Y H G F	1403			
<input checked="" type="checkbox"/> <a href="#">XP_007048682</a>	1329	L S L S L L I F E D D - H I T F I ----- P G N K p r N F S L I T P F S A D V W V I R I P S E T E S F S A R S S D T C I M A R I G I C Q V F V D D F Y F I G G F	1403			
<input checked="" type="checkbox"/> <a href="#">XP_006429807</a>	705	L I L S F V I L K D G G Y G C F L G E Q D P G N -- r N I I L M A P V S A D V W V R I P W E D K S N N E G S I A S T C I M S R I Q N C Q I I V D D C Y A Y H G F	782			
<input checked="" type="checkbox"/> <a href="#">XP_004305785</a>	1356	L F L S L L S F K D N Q Y S ----- S S I -- N V P L I G P L C A D I W E I P C E N E S S C Q S P S N T C V M I R V G N C Q L K P E D D H F F Q G F	1425			
<input checked="" type="checkbox"/> <a href="#">XP_002275536</a>	1914	L S L S L L I F K D A H D I L L M F G Q D S A P G -- N I T F I A P L S V D W V V R I P W E S E T L N G C S P A P M C V M V R V C N C Q L I A E D G Y I F S G F	1991			
<input checked="" type="checkbox"/> <a href="#">EXB26144</a>	1169	V F L S F L S F K D -- H G C L S L D K D A N C A -- D V T L I A P L S A D I W V R L P C E S E S S R K S T P L T T C I M S R I S E C Q V L A E D D H F F F G F	1244			
<input checked="" type="checkbox"/> <a href="#">XP_007217664</a>	469	L F L S F L S L K D D G R S Y L K F D D I E H V -- D I P L V A P L C A D W V V K I P C E N E S S -- S S P S T I C V M T R I K N C Q L M A E D A Q F F H G F	544			
<input checked="" type="checkbox"/> <a href="#">XP_002518393</a>	1324	L F L S L L I C K D D G Y G C L I L N E D N G F -- n N I T L I A P L S A D V W V R L P C E S P C L N S S S A S T C V M S R I A N C Q L H A D D C Y T L D G F	1401			
<input checked="" type="checkbox"/> 10001	359	----- Y L D D G H G C S M F D Q D T D S V -- Y V S L V A P L S A D V W V T I P C E S E S I L C R S S P S I T C V M T R I R D C Q I M A D D G H F F N G F	429			
<input checked="" type="checkbox"/> <a href="#">XP_006429201</a>	1404	D A L L D V I N Q F S S V N D E S K L F T C D V Q Q F L L K R C R R E N G A V S V V A S D T I F I D L R F C V D S I M I K L H R L R R D S G S L K P V A K L N	1483			
<input checked="" type="checkbox"/> <a href="#">XP_007048682</a>	1404	E A L L E I I D L F S V F Q D E S K Y M S D V L Q F L Q S K R K L R K E R A V S L L D S A M T F T E V R C Y V E S L L I Q L N R L R D S G S L K P V A K L N	1483			
<input checked="" type="checkbox"/> <a href="#">XP_006429807</a>	783	D A L L D V I N Q F S S V N D E S K L F T C D V Q Q F L L K R C R R E N G A V S V V A S D T I F I D L R F C V D S I M I K L H R L R R D S G S L K P V A K L N	862			
<input checked="" type="checkbox"/> <a href="#">XP_004305785</a>	1426	Q G I T D I I N Q F S I V S D L S E C F K T D V L Q F L Q S K R C L A Q N N E D P P V L S S V N Y T E V R C Y V N S L S I Q L N P C Q R N S -- E D P I A T E	1503			
<input checked="" type="checkbox"/> <a href="#">XP_002275536</a>	1992	E A L I D V F Q F S S I D E E S K C F T S D V L Q F L H S K R C L S R E S R A V P S K A S N M F T E A R C F V N S L S I K F C C I K D P S I S F E P V A K A D	2071			
<input checked="" type="checkbox"/> <a href="#">EXB26144</a>	1245	E A L I D V M N O F S I L V P D Q S C F K S D V P E F I O L K R C F K O N S V A S I A S V V I L T E V R C H V N S I L V M K F H O F T K G S -- T R I T A K A E	1322			

The prediction model is missing 54 amino acids between #277-278 and 28 between #358-359.

# COBALT data

- 54 missing residues between #277-278
- 28 missing between #358-359
- 42 missing between #683-684
- 4 missing between #689-690
- 5 missing between #2230-2231
- 22 missing between #2057-58
- 11 extra residues from #688-699
- 16 extra residues from #702-718

# Improved COBALT data

COBALT: – Cobalt RID ( seqs)

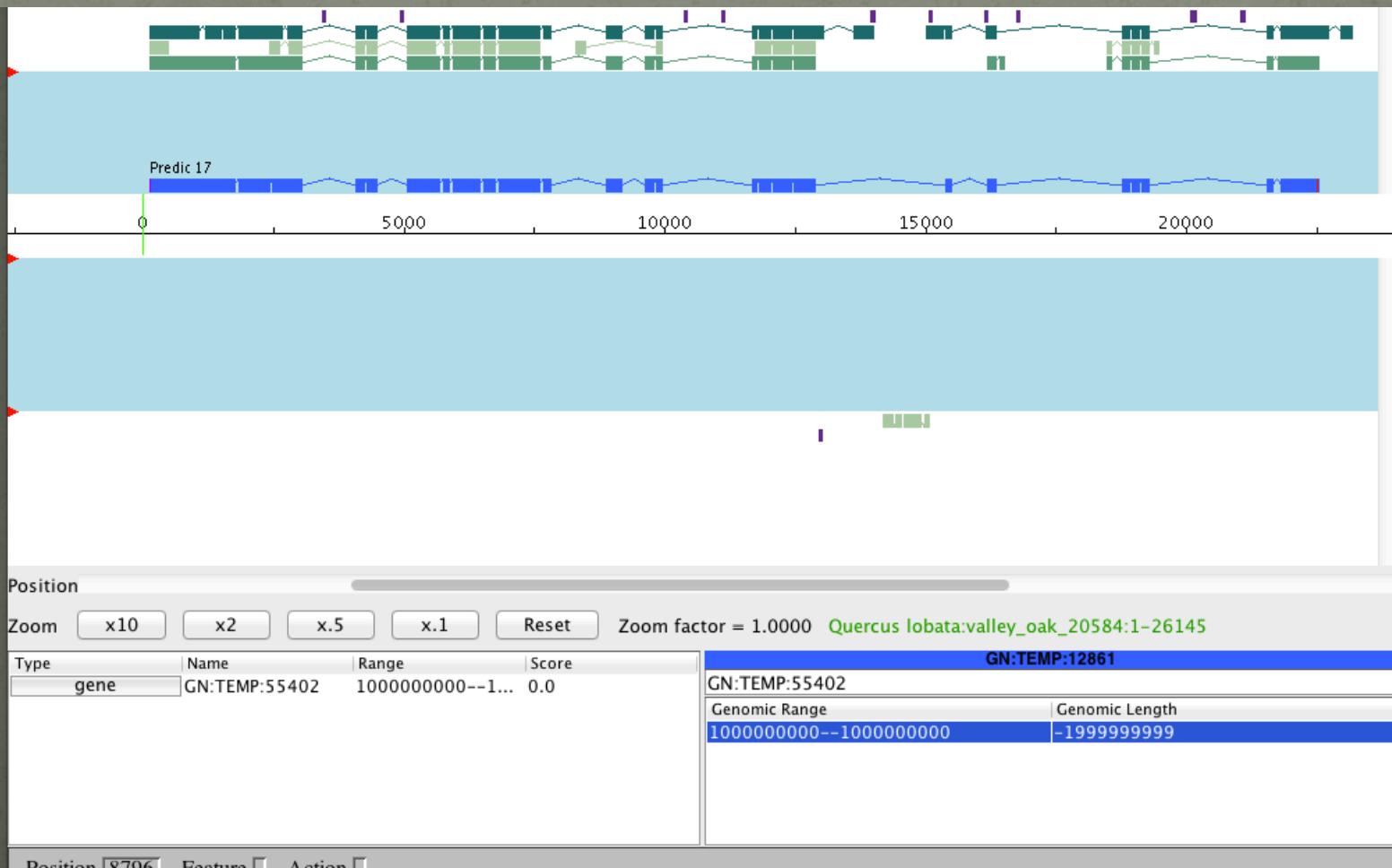
www.ncbi.nlm.nih.gov/tools/cobalt/cobalt.cgi

DNA Subway    www.fs.fed.us/psw/publications/...    NCBI Blast:Predict 17 (amino acid...)

COBALT: – Cobalt RID ( seqs)

<input checked="" type="checkbox"/> <a href="#">XP_007217664</a>	1673	KDVCYKQKGTDVFHIGIGEHSHLHWMDTAMELLVSIRYDEPGWQWSGGFLPDHLDGTQVKMRNYLSGSLNMRIVEVQNA	1752		
<input checked="" type="checkbox"/> <a href="#">EXB26144</a>	2412	KDLCFKQKGTDHFRLRMGEHSHLHWMDTTRELLSVRYNEPGWQWSGSFLPDHLDGTQVKMRNYVSGSSSVIRVEMQNA	2491		
<input checked="" type="checkbox"/> <a href="#">XP_007048682</a>	2591	KDIYYKQKGTDIVYHLCVGQHSQLHWTDTTRELLISMLFDEPGWQWSGSFLPDHLDGTQVKTRNYASGAMNMIRVEVQNA	2670		
<input checked="" type="checkbox"/> <a href="#">XP_006370741</a>	1785	KKICYKQKGTDVSRLGIGQHHHLHWKDTTRELLVSICFDEPGWEWSGSFLPDHLDGTQVKMRNN-AGVLRMIRVEVQNA	1863		
<input checked="" type="checkbox"/> <a href="#">XP_006594305</a>	2186	KEISYKQKGTDAVFYLGIGKHDHLHWHTDTTRELLVSICYNESGWQWSGSFLPDHLDGTQLKMRNYVFGTSNMRIVEVQNA	2265		
<input checked="" type="checkbox"/> <a href="#">XP_006594302</a>	2570	KEISYKQKGTDAVFYLGIGKHDHLHWHTDTTRELLVSICYNESGWQWSGSFLPDHLDGTQLKMRNYVFGTSNMRIVEVQNA	2649		
<input checked="" type="checkbox"/> <a href="#">XP_006594304</a>	2541	KEISYKQKGTDAVFYLGIGKHDHLHWHTDTTRELLVSICYNESGWQWSGSFLPDHLDGTQLKMRNYVFGTSNMRIVEVQNA	2620		
<input checked="" type="checkbox"/> <a href="#">XP_006594303</a>	2570	KEISYKQKGTDAVFYLGIGKHDHLHWHTDTTRELLVSICYNESGWQWSGSFLPDHLDGTQLKMRNYVFGTSNMRIVEVQNA	2649		
<input checked="" type="checkbox"/> 10001	1710	KDLCYKQKGTDVFQLGVGEAHLLHWHTDTTRELLSVRYNEPGWQWSGSFLPDHLDGTQVKLNRNVSNALNMVRVEVQNA	1789		
<input checked="" type="checkbox"/> <a href="#">XP_007217664</a>	1753	DVSMGDEKIVGNFHGNSGTNLILISDDETGYMPYRIDNFNSNERLRIYQQRCETVETTVHSYTSCPYAWDEPCYPHRLTVE	1832		
<input checked="" type="checkbox"/> <a href="#">EXB26144</a>	2492	DVSVRDEKVVGSLHGDSGTMILLSSDDDTGYMPYKIDNFSKERLRFQQKCDFETIVHSYTSCPYAWDEPCYPHRLTVE	2571		
<input checked="" type="checkbox"/> <a href="#">XP_007048682</a>	2671	DVSVRDE-IVGSLQGSSGTNLILLSEDDDTGYMPYRIDNFNSKERLRIYQQRCESLDTIVHPYTPSCPYAWDEPYYPHRVTIE	2749		
<input checked="" type="checkbox"/> <a href="#">XP_006370741</a>	1864	NVSVKDEKIIIGSLHGNSGTNLILLSSDDDTGFMPYRIDNFNSKERLRYQQKCENFDVTIHPYTPSCPYAWDEPCFPHRLTVE	1943		
<input checked="" type="checkbox"/> <a href="#">XP_006594305</a>	2266	DISMGDEKIVGNIKGNSGTNLILLSSDDDTGYMPYRIDNFNSKERLRIYQQRCEMFDVTIHSYTSCPYTWDPCYPRRLLIVE	2345		
<input checked="" type="checkbox"/> <a href="#">XP_006594302</a>	2650	DISMGDEKIVGNIKGNSGTNLILLSSDDDTGYMPYRIDNFNSKERLRIYQQRCEMFDVTIHSYTSCPYTWDPCYPRRLLIVE	2729		
<input checked="" type="checkbox"/> <a href="#">XP_006594304</a>	2621	DISMGDEKIVGNIKGNSGTNLILLSSDDDTGYMPYRIDNFNSKERLRIYQQRCEMFDVTIHSYTSCPYTWDPCYPRRLLIVE	2700		
<input checked="" type="checkbox"/> <a href="#">XP_006594303</a>	2650	DISMGDEKIVGNIKGNSGTNLILLSSDDDTGYMPYRIDNFNSKERLRIYQQRCEMFDVTIHSYTSCPYTWDPCYPRRLLIVE	2729		
<input checked="" type="checkbox"/> 10001	1790	DVSIQDEKVVGSLHGNSGTNLILLSSDDDTGYMPYRIDNFNSKERLRIYQQKCETFETIVHSYTSCPYAWDEPCYPHRLTVE	1869		
<input checked="" type="checkbox"/> <a href="#">XP_007217664</a>	1833	VPGKRVLGSYTLDVVKEYSPVQLPSSSEKRERTLHLSIHAEGATKVLHVIDSSYHILNDMKKTSVPRLREKRNDEQKQDK	1912		
<input checked="" type="checkbox"/> <a href="#">EXB26144</a>	2572	VPGERVLGSYSLDEVKEYIPVLDLPPSSEKPRKLVLSVHAEGATKVLRVIDSNYHILNDTENSSGPYLREKKQEQKQDK	2651		
<input checked="" type="checkbox"/> <a href="#">XP_007048682</a>	2750	VPGERIVGFSFLDDLKEYMPVHLQSTSEKPERMILLSVRAEGATKVLSIDSTYHILKDMEDHSTIRFQEKKQEQEKQEQK	2829		
<input checked="" type="checkbox"/> <a href="#">XP_006370741</a>	1944	VPGQRIVIGSYALDDLKEYIPVQLKATAEKPRTLILSVHAEGATKVLGIVDSSFHVLKDVKDPSPPWFREKTKHEQKQDK	2023		
<input checked="" type="checkbox"/> <a href="#">XP_006594305</a>	2346	VPGERVLGSYDLDDVKEYPVYLPSTSEKPARTFYLVSVAEGATKVLVSLVDSNYHIFNDVKKSSVPLPTEKRLCDHSLVR	2425		
<input checked="" type="checkbox"/> <a href="#">XP_006594302</a>	2730	VPGERVLGSYDLDDVKEYPVYLPSTSEKPARTFYLVSVAEGATKVLVSLVDSNYHIFNDVKKSSVPLPTEKRLCDHSLVR	2809		
<input checked="" type="checkbox"/> <a href="#">XP_006594304</a>	2701	VPGERVLGSYDLDDVKEYPVYLPSTSEKPARTFYLVSVAEGATKVLVSLVDSNYHIFNDVKKSSVPLPTEKRLCDHSLVR	2780		
<input checked="" type="checkbox"/> <a href="#">XP_006594303</a>	2730	VPGERVLGSYDLDDVKEYPVYLPSTSEKPARTFYLVSVAEGATKVLVSLVDSNYHIFNDVKKSSVPLPTEKRLCDHSLVR	2809		
<input checked="" type="checkbox"/> 10001	1870	VLGERVLGSYALDDVKDYMPVYLPSTSEKPERTLRLSVNAEGATKVLVIDSSFHILNDMKNSSAHLREKGKHAQKQDK	1949		

# Best Consensus Model



# Functional Analysis: InterProScan 5 Results

Similar proteins  
Structures

Filter view on

Entry type

F Family

D Domains

R Repeats

S Site

Status

Unintegrated

Colour by [help](#)

domain relationship

source database

## PREDIC

Length 2,593 amino acids

### Protein family membership

Vacuolar protein sorting-associated protein 13 (IPR026847)

### Domains and repeats

1 500 1000 1500 2000 2593 ► Domain

### Detailed signature matches

IPR026847 Vacuolar protein sorting-associated protein 13

► PTHR16166 (VACUOLAR...)

IPR009543 Vacuolar protein sorting-associated protein 13 domain

► PF06650 (DUF1162)

no IPR Unintegrated signatures

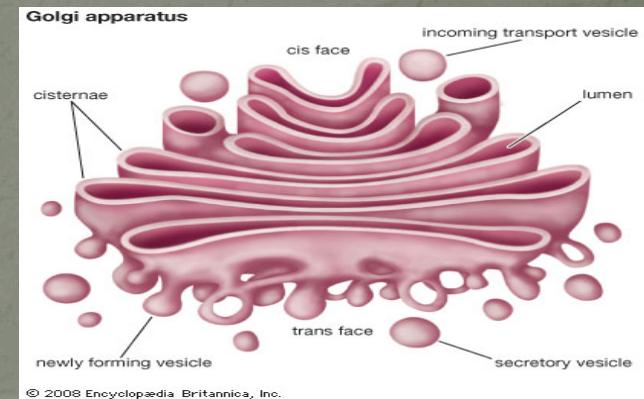
► PTHR16166:SF86 (SUB...)

### GO term prediction

Biological Process

# Vacuolar Protein Sorting-Associated Protein Family 13

- Play a role in controlling the proteins cycling through the trans-Golgi network.
- ATPase required for endosomal trafficking.
- Regulates membrane-bending activity that generates intraluminal vesicles.
- Mutations in humans cause the debilitating hereditary neurodegenerative diseases chorea acanthocytosis and Cohen syndrome.

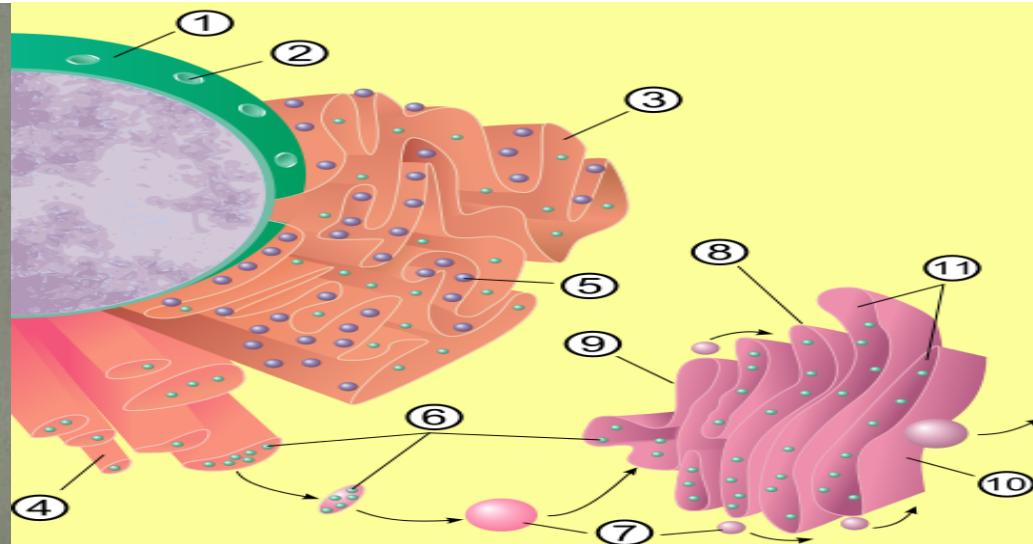


# Localization Analysis: PSORT Results

----- Final Results -----

```
plasma membrane --- Certainty= 0.600(Affirmative) < succ>
Golgi body --- Certainty= 0.400(Affirmative) < succ>
endoplasmic reticulum (membrane) --- Certainty= 0.300(Affirmative) < succ>
microbody (peroxisome) --- Certainty= 0.300(Affirmative) < succ>
```

----- The End -----

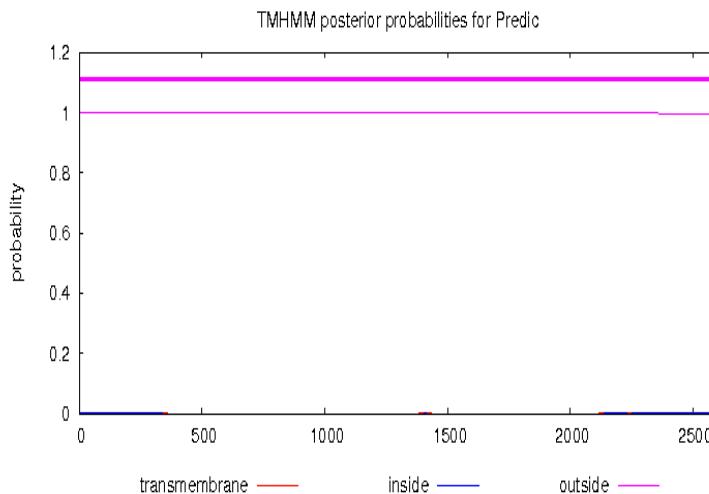


# Localization Analysis: TMHMM      Phobius

## TMHMM result

[HELP](#) with output formats

```
# Predic Length: 2593
# Predic Number of predicted TMHs: 0
# Predic Exp number of AAs in TMHs: 0.1219
# Predic Exp number, first 60 AAs: 0.00049
# Predic Total prob of N-in: 0.00146
Predic TMHMM2.0      outside    1 2593
```

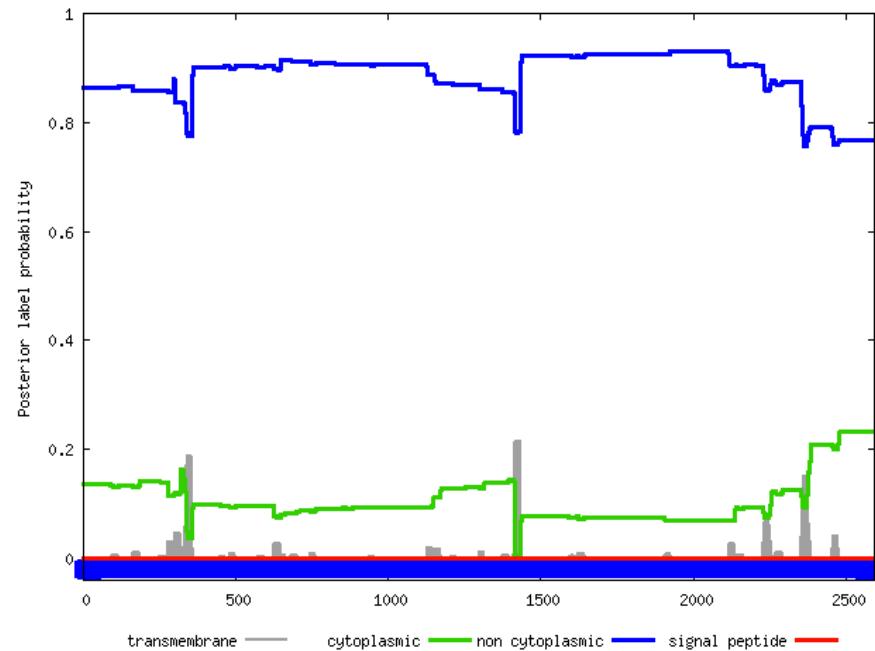


# [plot](#) in postscript, [script](#) for making the plot in gnuplot, [data](#) for plot

## Predic

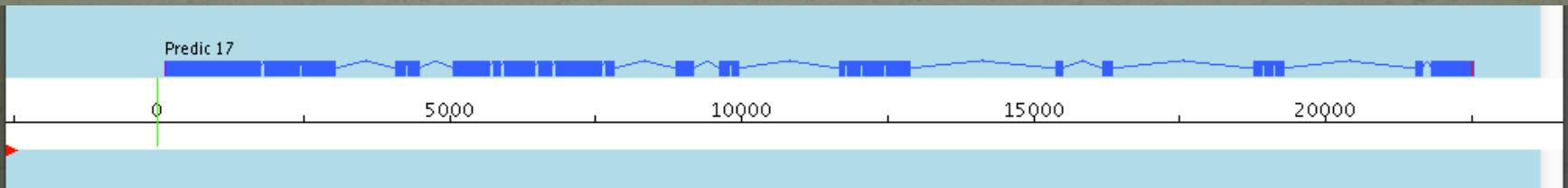
ID	Predic	FT	TOPO_DOM	1	2593	NON CYTOPLASMIC.
//						

Phobius posterior probabilities for Predic



The probability data used in the plot is found [here](#), and the gnuplot script is [here](#).

# Conclusions: Contig 18000137



- Localized within the lumen of the ER, golgi complex, or within the vacuoles and peroxisomes it packages.
- Codes for a protein belonging to vacuolar protein sorting-associated protein family 13.
- Retaining a nonmutated copy of this gene may be crucial for the oak's survival.

# Conclusions



- The fully annotated contig 18000137 can be added to the growing reference genome for *Quercus lobata*.
- This genome can power future conservation strategies employed to ensure that the common valley oak, a foundational species, will enjoy prosperity with its maximum evolutionary potential intact.

End