

**Appendix D: Completed Annotation Report for the *Spinophilin G* Isoform of *Drosophila erecta***

**Annotation report**

Student Name: xxxxxxxx & xxxxxxxxxx  
Student E-mail: xxxxxxxx@amherst.edu & xxxxxxx@amherst.edu  
Faculty Advisor: Dr. Julie Emerson  
College/University: Amherst College

Project name: d erecta\_2<sup>nd</sup>3Lcontrol\_Nov2011\_fosmid52  
Project species: *Drosophila erecta*  
Date of submission: Dec. 7, 2011  
Size of project in base pairs: 28069  
Number of genes in project: 1  
Complete Report

For each gene complete the following Gene Report Form (copy and paste to create as many copies as needed, be sure to create enough isoform reports within your gene form for all isoforms):

=====Gene Report Form=====

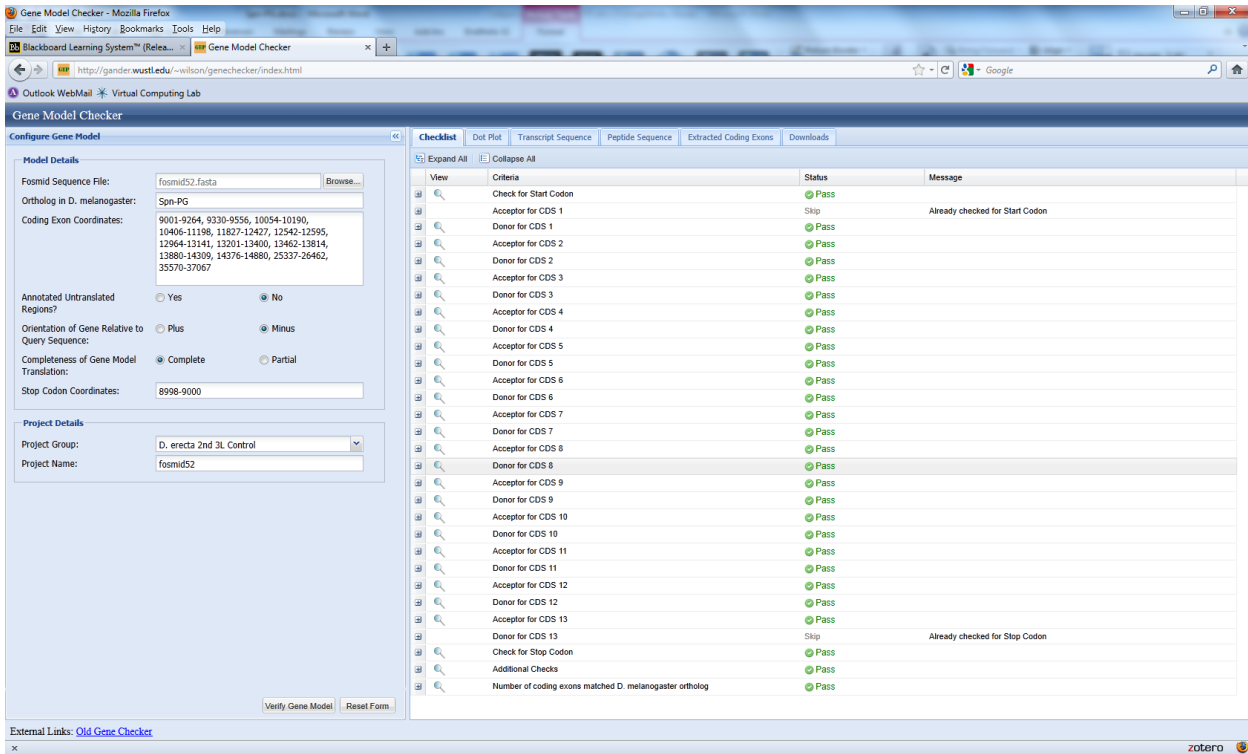
Gene name: D erecta Spinophilin  
Gene symbol: dere\_Spn  
Approximate location in project (from start codon to stop codon): 37067-8998  
Number of isoforms in *D. melanogaster*: 10  
Number of unique isoforms based on coding sequence: 8  
List names of unique isoforms (i.e. PA, PC etc): PB/PC/PH, PI, PJ, PG, PF, PD, PE, PK  
Number of unique isoforms found in this project: 1

=====Isoform report=====

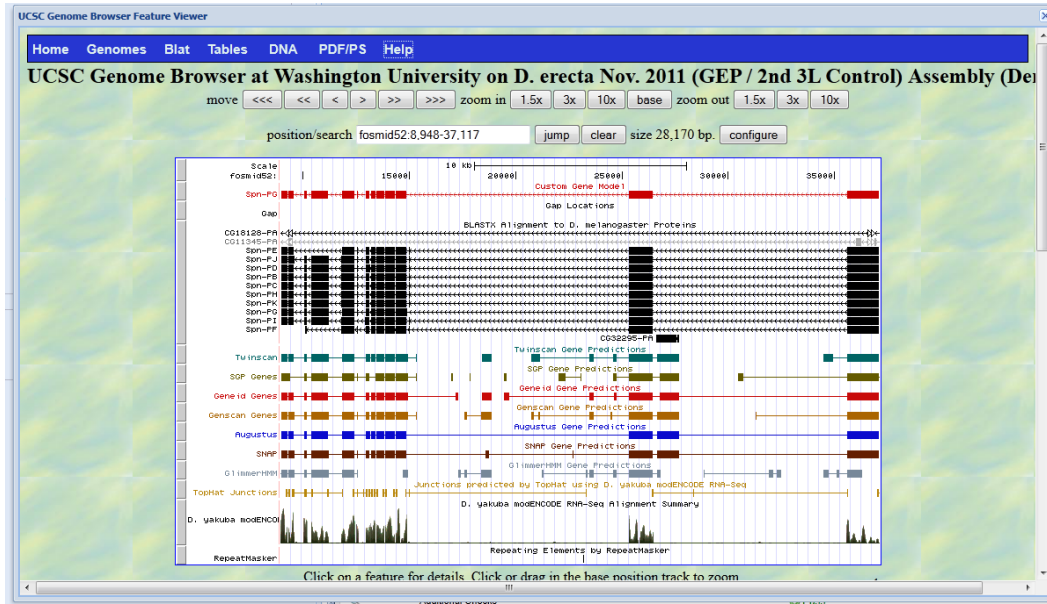
For each (protein) isoform complete the following (copy and paste to create as many copies as needed):

Gene-isoform name: dere\_Spn-PG  
Is the 5' end of this isoform missing off the end of project: no  
Is the 3' end of this isoform missing off the end of the project: no

Enter the coordinates of your final gene model for this isoform into the gene model checker and paste a screen shot of the results below:



Using the custom track feature from the Gene Model Checker, capture a screen shot of your gene model shown on the browser for your project; zoom in on only your isoform. **If available**, also show these tracks: other ref seq; all relevant Gene Prediction Tracks, 3-way and 5-way multi Z). If you need help, see lab instructor and/or read the bottom of page 9 in The Gene Model Checker User Guide, on the “Documentations” page under the Help menu at gep.wustl.edu. (Type comments about your model below the screen shot.):



Each of the gene predictor tracks is very consistent with our model. As can be seen from the screen shot there are no major differences between my model of the Spn-PG gene and the various gene predictor tracks, thus supporting our model.

Do an alignment using blast2seq of the predicted gene model protein coding sequence compared to the protein sequence from *D. melanogaster*. Copy and paste the complete results below. (Type comments about the quality of the alignment below it.):

```
>lcl|50093 unnamed protein product
Length=2122

Score = 3671 bits (9520), Expect = 0.0, Method: Compositional matrix adjust.
Identities = 2047/2102 (97%), Positives = 2072/2102 (99%), Gaps = 5/2102 (0%)

Query 1      MEKPMHHAPAPVGKVSQIANIFQRKPIEIQPVEQLSAVAAAHAAAAAAAAAAHHAHVQGA 60
Sbjct 1      MEKPMHHAPAPVGKVSQIANIFQRKPIEIQPVEQLSAVAAAHAAAAAAAAAA H   QGA
MEKPMHHAPAPVGKVSQIANIFQRKPIEIQPVEQLSAVAAAHAAAAAAAAAAHHAH-AQGA 59

Query 61     PAVRTEHSARFNNARALFEKLGVESNSNVSSRLLRSGSREDNLCDGSDRSSRSDRSQ 120
Sbjct 60     PAVRTEHSARFNNARALFEKLGVESNSNVSSRLLRSGSREDNLCDGSDRSSRSDRSQ 119

Query 121    SPPKRRTPFPGVSLVHNNNNAAIVAQNGVPPEQRLSNSKFIVEPAAQVVPTSVMKYPQH 180
Sbjct 120    SPPKRRTPFPGVSLVHNNNNAA VAQNGV PEQRLSNSKFIVEPAAQVVPT+VMKYPQH 179

Query 181    NISRLKSEEPSVPPPPASGSVSALFASSGGDKPEKPERKFNSRELIEKQKKWTSHFTTKT 240
Sbjct 180    NISRLKS+EPSP+PPPASGSVSALFASSGGDKPEKPERKFNSRELIEKQKKWTSHFTTKT 239
```

Query	241	TTRTHSDLNRCDIIRTVPGTGLIMDSEKVAKPAMEPPQPPPNASPNPPMRAQAPPEIKPR	300
Sbjct	240	TTRTHSDLNRCDIIRTVPGTGLIMDSEKVAKPAMEPPQP PNASPNPPMRAQAPPEIKPR	299
Query	301	SGKIGSPVKSPLPPIPAVKPKNVSPVKFNPDRLRQSPTKTADNSPPPPPAKSAAVLQRS	360
Sbjct	300	SGKIGSPVKSPLPPIPAVKPKNVSPVK+NPDRLRQSPTKTADNSPPPPPAKSAAVLQRS	359
Query	361	LMQEQQELLRNSCDQGVAPIPPEKPRKKSVDLIEDTLPLTNCSTPSSCASPTSSYLMQPA	420
Sbjct	360	LMQEQQELLRNSCDQGVAPIPPEKPRKKSVDLIEDTLPLTNCSTPSSCASPTSSYLMQPA	419
Query	421	KRGSLDGGSGNGQYPGNLSGSTNSATSGSPVASASSGSSPVHTEDEKQENESTEKSEM	480
Sbjct	420	KRGSLDGGSGNGQYPGNLSGSTNSA SGSPVASASSGSSPVHTEDEKQENESTEKSE+	479
Query	481	EYYHGGNYNSVPRRRRSENEGRKSVDESSPSANNSQQQQQHSIPGSAAGSPQRVANKRSS	540
Sbjct	480	EYYHGGNYNSVPRRRRSENEGRKSVDESSPSANNSQQQQQHSIPGSA GSPQRVANKRSS	539
Query	541	ITVNMPAAGLGQRPPSIIISTTSQDEGGFNESAPELKAKLQPAYDQTEEQPHSLNYVDVGY	600
Sbjct	540	ITVNMPAAGLGQRPPSIIISTTSQDEGGFNESAPELKAKLQPAYDQTEEQPHSLNYVDVGY	599
Query	601	RLNPDGSESRVYVGSEAELYDTAKVTDQMQRKFGANGFGQESSTVYAI IKPDVQESQPVA	660
Sbjct	600	RLNPDGSESRVYVGSEAELYDTAKVTDQMQRKFGANGFGQESSTVYAI IKPD+QESQPVA	659
Query	661	PSRSVLIQSPNSSSVEGSPLHRGSYSSPPVGVVSPIRRNRSSNQDQSVGGGG--SAKTTP	718
Sbjct	660	P+R VL+QSP SSSVEGSPLHRGSY+SPPVGVVSPIRRNRSSNQDQ VGGGG SAK+TP	719
Query	719	QCSPARSALVKGIAPIASIDAHEEEELDLVEEDEHLAVEYVEVLELQDDEEEEEEAPVLPE	778
Sbjct	720	PCSPARSAMVKGIAPIASIDAHEEEELDLVEEDEHLAVEYVEVLELQDDDEEEEAPVLPE	779
Query	779	RRAPAQGSLELQDLEYADTSAGEDEEDI INHLKDGVDLDELIDDVVDEVIKVHVNHNSVA	838
Sbjct	780	RRAPAQGSLELQDLEYADTSAGEDEEDI INHLKDG+LDVELIDDVVDEVIKVHVNHNSVA	839
Query	839	TAPSIQAATPAAAI PREDSLPDDMTAAEAERLLSSRQQLSLLSDEQAKEVEQILNAAPSVG	898
Sbjct	840	TAP IQAA PAAAI PR DSLPDDMTAAEAERLLSSRQQLSLLSDEQAKEVEQILNAAPSVG	899
Query	899	VAVATVVATATSPTS IKNLIEDLPGQSAVAASAANGEQDIQIAAVPAIVEEDEDEEEDFP	958
Sbjct	900	VAVATVVATATSPTS IKNLIEDLPGQ+AVAASAANGEQDIQIAAVPAIVEEDEDEEEE+	959
Query	959	EDDEED-HARADFDANGGDADGSDSDDVEAVDIVGYGHASTALNATFVKADSTETETTTTT	1017
Sbjct	960	++D+E HARADFDANGGDADGSDSDDVEAVDIVGYGHASTALNATFVKADSTETETTTTT	1019
Query	1018	PSTATTATTRHDDDEPEWLRDVLEAPKRSLENLLITSATSSRAPGQREELNGYDLHEKH	1077
Sbjct	1020	PSTATTATTRHDDDEPEWLRDVLEAPKRSLENLLITSATSSRA GQREELNGYDL EKH	1079

Query	1078	SDLNQTYITGGESLHESIVSVESTQSDATLNQTTTIDDSIISSKHNSTYSLADAEQATSS	1137
Sbjct	1080	SDLNQTY+TGGESLHESIVSVESTQSDATLNQTTTIDDSIISSKHNSTYSLADAEQAT+S	1139
Query	1138	TVLSTGVTELDSDSQQYYIPEYPPVRSKEVLVEAGVHYFEDGNFWMEVPGLLDFDDDDCSYP	1197
Sbjct	1140	TVLSTGVTELDSDSQQYYIPEYPPVRSKEVLVEAGVHYFEDGNFWMEVPGLLDFDDDDCSYP	1199
Query	1198	PITVRKNPKVRFSSGPIHVYSTFSVNDYDRRNEDVDPVAASAEYELEKRVEKMHVFPVEL	1257
Sbjct	1200	PITVRKNPKVRFSSGPIHVYSTFSVNDYDRRNEDVDPVAASAEYELEKRVEKMHVFPVEL	1259
Query	1258	MKGPEGLGLSIIIGMGVGADAGLEKLGIFVKTTITDNGAAARDGRIQVNDQIIIEVDGKSLVG	1317
Sbjct	1260	MKGPEGLGLSIIIGMGVGADAGLEKLGIFVKTTITDNGAAARDGRIQVNDQIIIEVDGKSLVG	1319
Query	1318	VTQAYAASVLRNTSGLVKFQIGRERDPENSEVAQLIRLSLQADREKEERLKRQQEEYLRR	1377
Sbjct	1320	VTQAYAASVLRNTSGLVKFQIGRERDPENSEVAQLIRLSLQADREKEERLKRQQEEYLRR	1379
Query	1378	TLDYSEDSTQPVANSVCEGPSSPVQVEHPMEVEATHSQEVESLKRLQLQESMGCLVKE	1437
Sbjct	1380	TLDYSEDSTQPVANSVCEGPSSPVQVEHPMEVEATHSQEVESLKRLQLQESMGCLVKE	1439
Query	1438	EIIQNLKRKLVKLETTGNENELLSERLRQSERELGNIRKEAANLQNMLQSSQGYMALDK	1497
Sbjct	1440	EIIQNLKRKLVKLETTGNENELLSERLRQSERELGNIRKEAANLQNMLQSSQGYMALDK	1499
Query	1498	KYNKAKRLVREYQQRELDMCHREEFYQQLLQEKDTEYNALVKKLKDRVINLEHELQETQR	1557
Sbjct	1500	KYNKAKRLVREYQQRELDMCHREEFYQQLLQEKDTEYNALVKKLKDRVINLEHELQETQR	1559
Query	1558	KAGFPVGLPYDSATLKLTPQMMRKTTPKPLFHKLETELSDTEISDLSPDGDGVTATVER	1617
Sbjct	1560	KAGFPVGLPYDSATLKLTPQMMRKTTPKPLFHKLETELSDTEISDLSPDGDGVTATVER	1619
Query	1618	KVPVKDELDAAVPQHELLDNSINKTKIDLNRQLPSANGNSSTSNAAVDLQLSNGNLL	1677
Sbjct	1620	KVPVKDELDAAVPQHELLDNS+NKTKIDLNRQLPSANGNSSTSNAAVDLQLSNGNLL	1679
Query	1678	KRSRNSRSDCTLDDEEEERESEALNLAGAPVAHETISLSNGNSHLLANVNNLLQHH	1737
Sbjct	1680	KRSRNSRSDCTLDDEEEERESEALNLAG PV HETISLSNGNSHLLANVNNLLQHH	1739
Query	1738	PPAMATVIATPSNGHLGTTTTPILLNSTSSASSSSSNQSTAREAQINQLYAQVHKDPSKQQ	1797
Sbjct	1740	PPAMA+V+ATPSNGHLGTTTTPILLNSTSSASSSSSNQSTAREAQINQLYAQVHKDPSKQQ	1799
Query	1798	HQQQQQQQQAQAVTTSIPSIKFNALGSPADNGLNDFHRGSMTTFGTGPATSSNRDLNSS	1857
Sbjct	1800	-HQQQQQQQQAQAVTTSIPSIKFNALGSPADNGLNDFHRGSMTTFGTGPATSSNRDLNSS	1858
Query	1858	YDSILGSNDKLAENDPAESWMPYPSRRRVAPNGSKVPLPGSSFTDQLNQALSDRERRLGDG	1917
Sbjct	1859	YDSILGSNDKLAENDPAESWMPYPSRRRVAPNGSKVPLPGSSFTDQLNQALSDRERRLGDG	1918

Query	1918	SSRHSSDDYTEINKSQSAAAINCKTLN	IRQAVNEAQPKVPWQQQHHQQIQQQPSAHTT	1977
Sbjct	1919	SSRHSSDDYTEINKSQSAAAINCKTLN	IRQAVNEAQPKVPWQQQHHQQIQQQPSAHTT	1978
Query	1978	GPPSPTSMSSGCSSPGYSPSRTL	DLGSSSSFSDRKAMAAGYTYKGGPVHEWTKDQVGHW	2037
Sbjct	1979	GPPSPTSMSSGCSSPGYSPSRTL	DLGSSSSFSDRKAMAAGYTYKGGPVHEWTKDQVGHW	2038
Query	2038	LMGIELERYIPVFKENNVEGGALL	TLDSKDFKTLGICGDDKHRLKKRLKDLKANIEKERK	2097
Sbjct	2039	LMGIELERYIPVFKENNVEGGALL	TLDSKDFKTLG+CGDDKHRLKKRLKDLKANIEKERK	2098
Query	2098	DM	2099	
		DM		
Sbjct	2099	DM	2100	

As can be seen from the excellent correlation between our predicted protein in *D. erecta* against the equivalent protein in *D. melanogaster* in the Blastp, our prediction is most likely very accurate.